# MEASUREMENT TO INTELLIGENCE: FEATURE EXTRACTION, MODELING AND PREDICTIVE ANALYSIS OF ASYMMETRIC CONFLICT EVENTS

A Dissertation

by

STEPHEN MICHAEL GEORGE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Jyh-Charn Liu |
| Committee Members, | Riccardo Bettati |
| | Guofei Gu |
| | Michael Bishop |
| Head of Department, | Nancy M. Amato |

August 2014

Major Subject: Computer Science

# ABSTRACT

The conflict events that comprise asymmetric warfare are a primary killer of both combatants and civilians on the modern battlefield. Improvised explosive devices (IED) and direct fire (DF), the most common of these attacks, claim thousands of lives as conventional and unconventional forces clash. Computer-based predictive analysis can be used to identify locations that are useful for these events, potentially providing the awareness needed to disrupt or avoid attacks before they are launched.

In this dissertation, I propose an analytical framework for predictive analysis of asymmetric conflict events. This framework incorporates a tactics-aware system model based on attacker roles that is populated with a set of geomorphometric and visibility-constrained features describing terrain and proximity to necessary supporting structures. Features that identify and assess the utility of terrain for use by risk-averse attackers are important contributors to the model. Statistical learning is used to extract spatially and temporally constrained tactical patterns. These patterns are then used to predict the utility of future or unvisited locations for conflict events.

Major contributions of this dissertation include:

(1) A concise, accurate feature representation of conflict events in non-urban environments;

(2) A system model based on attacker roles that captures the tactical patterns of conflict events;

(3) Accurate conflict event classification algorithms that support predictive analysis; and

(4) A novel method for detecting and describing features that support risk-averse attackers.

The framework has been implemented and tested on real-world IED and DF data collected from the conflict in Afghanistan in 2011-2012. Several learning techniques are assessed using two dimensionality reduction schemes under a variety of spatial, temporal and combined constraints. A resource-unconstrained version of the framework accurately predicts conflict events across a wide range of terrain types and over the 19 months covered by available data. A limited version of the framework that assumes less computational capability provides useful predictive analysis that can be performed in mobile and resource constrained environments.

# DEDICATION

To Ana

# ACKNOWLEDGEMENTS

Research is seldom the work of a single person. This project has benefitted from the contributions of many people, both directly and indirectly. I would like to especially acknowledge the following people.

First and foremost, I want to thank my wife Ana for her encouragement and faith in me. I also thank my daughter Anita for understanding the early bedtimes and study weekends. Thanks to my parents, Dr. William T. George and Mrs. Nelda A. George, for instilling a strong work ethic and a love of reading and learning.

I would like to express my gratitude to my advisor Dr. Jyh-Charn Liu. He is a strong leader and thoughtful mentor. This research project was only possible thanks to his patience, teaching, and willingness to see the world through the eyes of a soldier. We were brought together by the late Dr. William 'Mac' Lively, a professor who positively influenced my life and the lives of generations of Texas A&M students.

Two of my fellow students were key contributors. I owe a debt to U.S. Air Force Maj. John Pecarina, PhD, for his assistance at a particularly critical time in my academic career. Without John's patience and help, I might have failed before this project truly started. My thanks also go to U.S. Army LTC Tim Nix, PhD, for thoughtful conservation, brainstorming sessions and, maybe most importantly, for providing a pointer to some critical data.

For 21 years, the United States Army provided me with a rich set of opportunities and experiences. It also provided me with a second family. Chris 'Gump' Graves, my

fraternal twin, worked alongside me for 11 years. His wise counsel and excellent technical skills helped me lay the foundations that underpin this work. I would also like to acknowledge Pablito B., Chris S., Charlie C., Bob A., Mike G., Alex P., Nick L., George P. and Rick S. You all played important roles over the years. To Mike, Robb, Oscar and Buzz, I appreciate the conversations and laughter. A special thanks to all of my brothers and sisters in the Armed Services who mentored and led me over the years.

*Send me*

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

# CHAPTER I

# INTRODUCTION

Low intensity conflict (LIC) is defined by the U.S. military as "a political-military confrontation between contending states or groups below conventional war and above the routine, peaceful competition among states." [1]. These confrontations can range from organized protest to armed combat and use a variety of social, political, and economic tactics. Similar conflict exists between criminals and law enforcement. In both types of conflict, opposing sides attempt to gain advantage by exploiting *asymmetry*. For military conflicts, there is often asymmetry of force: a large number of attackers engage in combat with a smaller number of targets. The classic ambush is an example. Asymmetry may also be kinetic, where attackers employ some weapon that exploits or maximizes target vulnerabilities. Improvised explosive devices (IED) leverage this type of asymmetry. For both military and criminal activities, there may be asymmetry of information. The burglar is likely to know more about neighborhood patterns than the policeman. Similarly, an attacker from the local area may take advantage of knowledge of roads and paths to outmaneuver a target from a distant location.

An asymmetric conflict event -more simply a 'conflict event'- is a particular instance of the cat-and-mouse game played between opposing forces with unequal capabilities or knowledge. Collectively described as asymmetric warfare by the military, conflict events include both lethal and non-lethal encounters between forces where the encounter is planned and executed in order to obtain some specific outcome like

casualties or favorable propaganda. A key concept of this definition is the idea that all elements of the conflict event are selected by the attacker using unknown and possible unknowable criteria. Thus, the choices made by the human decision maker are critical to the execution of the event.

In many conflict events, the impact of correct decisions is clear. For example, asymmetry of force is the result of human preparation and planning where attack location and configuration are chosen to ensure a minimum number of targets are engaged by a maximum number of attackers. A human chooses to initiate the event only when the attacking force is large enough to dominate the target. Similarly, a command-detonated IED is carefully emplaced in a location that provides adequate concealment along roads and paths that can be carefully monitored. The IED will only be triggered against a vulnerable target vehicle. Civilian conveyances and heavily armored transports are ignored. In both of these cases, the attack will occur only when a human assesses that the outcome will be favorable.

The exact choices made in the preparation and execution of a conflict event will almost always be unknown prior to the attack. In the typical case, there are many unknowns, most of which are difficult or impossible to measure:

- Number of attackers
- Available war material
- Attacker training
- Doctrine
- Change drivers
- Available tactics
- Past successes and failures
- Attacker experience
- Organization
- Desired outcome

Choices made regarding these unknowns are the results of complex, decision-making processes that are influenced by a variety of factors:

- Local terrain constrains attack emplacement site and execution strategies.
- Nearby terrain provides critical support elements for control and overwatch but may constrain the utility of otherwise satisfactory attack sites.
- War material and personnel availability/training constrain tactics.
- The sophistication of the target and its ability to react or respond with countermeasures influences tactical choices and drives attack evolution.

In spite of the unknown factors and influences, the execution of a conflict event exposes some of the decisions made by the attackers. For example, once the attack is complete, several things are known or observable: site and nature of the attack, number and location of active attackers, likely locations of observers and overwatch positions, and the advantage gained by the attacker from the terrain. Collection and analysis of these knowns allows a target to leverage its strengths to improve the predictability and recognition of high-threat locations. These strengths include:

- An ability to collect and retain measurements of historical conflict event sites.
- Knowledge of human factors in combat, acquired over many years of warfare against a variety of adversaries.
- Powerful computers and algorithms.

Thus, the main theme of this dissertation is the development of a conflict event model that incorporates measurable outcomes of the decisions made by attackers in the

preparation and execution of a conflict event. These results of these decisions, captured as features of the attack site and its environs, provide insight into an unknown and probably unknowable system that produces conflict events. This model –essentially a hidden human behavior model constrained by terrain and tactics– adaptively evolves to changing conflict situations. Two observations about human and their decision-making processes are central to its development.

First, humans instinctively avoid risk and seek benefit with their choices. This risk aversion is a factor of many conflict event choices regarding siting, timing and use of terrain. It is important to note that risk aversion is not necessarily correlated with fear. Risk aversion simply implies that risk is mitigated when possible. Similarly, benefits do not have to accrue to the decision-maker. The benefits may be abstract or may accrue to some other actor in the conflict event.

Also, all non-trivial human decision-making processes are probably unknowable. Humans integrate data collected by their senses, information from other humans, past experience, and a myriad of other inputs to make even simple decisions. Therefore, observable outcomes of a decision, like the emplacement location of an IED, are, at best, pale reflections and partial measurements of a complex and unobservable human process.

Based on these two observations, this work begins to interpret human behavior into math and programming and incorporates the result into a framework for modeling and predictive analysis of conflict events. Historical conflict events and their environs are modeled at various resolutions to capture the site-related choices made by the

humans involved. Beyond conventional terrain analysis, the model incorporates measurable characteristics of the actors' historical site selection and event execution processes and may be used to estimate the utility of unvisited sites for conflict events.

Modeling human behavior directly is difficult. Quantification of abstract concepts like experience, level of training, and risk tolerance is impossible without direct access to the individuals being modeled or similarly trained and motivated surrogates. In this research, we focus on capturing features that quantify observable results of behavior-driven choices. Risk aversion, human eyesight and rifle marksmanship are all used to constrain the range of locations suitable for the conflict event site and nearby support requirements. While the degree of risk aversion is impossible to determine from the existing dataset, it is possible to observe that features implying risk aversion are common to many conflict event sites. Likewise, many direct fire sites share common viewshed features, supporting the intuition that humans who shoot tend to select similar vantage points and fields of fire. While none of the features mentioned directly assess human behavior, they all reflect the conscious and unconscious human decision processes involved in their selection. The production of features that capture human behavior or the outcome of human decision-making processes is an important contribution of this research.

Complicating this modeling effort is the varying importance of features over terrain and time. A set of features that accurately captures a tactic and site used one year may completely fail to describe another tactic and site used one year later. Conflict events in rugged terrain may require a very specific geographic configuration while flat

terrain is less constraining. Different attackers with different experience and training may adopt different tactics, even in similar terrain. Even sets of events that share the same features may weight those features differently. Adaptively selecting and weighting relevant features to populate a realistic model is an important contribution of this research.

The anchoring of the features and associated model in real historical data is an important distinguishing feature of this research. The original conflict event data used for this analysis is messy. It is the outcome of real events in the real world where bullets were fired and explosions were triggered. Thus, there is no clean control set and no firm hypothesis against which all events can be tested. The quality of individual measurements is unknown and the manner in which key information was collected probably varied. Instead, the data emerges from a series of uncontrolled experiments (conflict events) where location and timing is estimated, the data collector and collection methods were different for every event, and the true desired outcome is unknown. Developing models and incorporating learning algorithms that can handle this type of uncertainty is an important contribution of this research.

The overarching purpose of this research project is the protection of human life. The bottom line is that targets with an ability to predict possible conflict event sites are more likely to survive an attack. As proposed in this research, predictive analysis can play an important role in conflict analytics by neutralizing knowledge asymmetry and providing situational awareness and decision-support facilities.

## Overview of Asymmetric Conflict Events

US military conflicts in the last 40 years have been largely asymmetric. Over this time and in years to come, the operational needs of asymmetric warfare (AW) are increasingly defined by the availability and delivery of timely and relevant information for decision makers and tactical forces. The cost of failure in terms of personnel, public opinion, money and equipment is a strong driving force behind developing counter-AW methods, yet AW challenges remain formidable because of changing threats and a high expectation of success.

Actors in these conflicts leverage asymmetry to achieve their goals. In counter-insurgency operations, the military enjoys overwhelming firepower and a high degree of mobility while the insurgents are more familiar with the terrain and have the tacit support of the civilian population. The military has sophisticated intelligence gathering abilities but may be blind in areas where its sensors can't reach while insurgents enjoy access to less sophisticated but more pervasive reports from the population that hosts them.

This dissertation examines two types of asymmetric conflict in Afghanistan: improvised explosive devices (IED) and direct fire attacks (DF).

### *Improvised Explosive Devices*

An IED is a homemade bomb constructed from materials easily available to people located in conflict areas. Its construction may range from an unsophisticated mix of common household cleaning products to extremely sophisticated assemblies involving custom circuit boards, shaped charges and remote triggers. Most IEDs lie between these

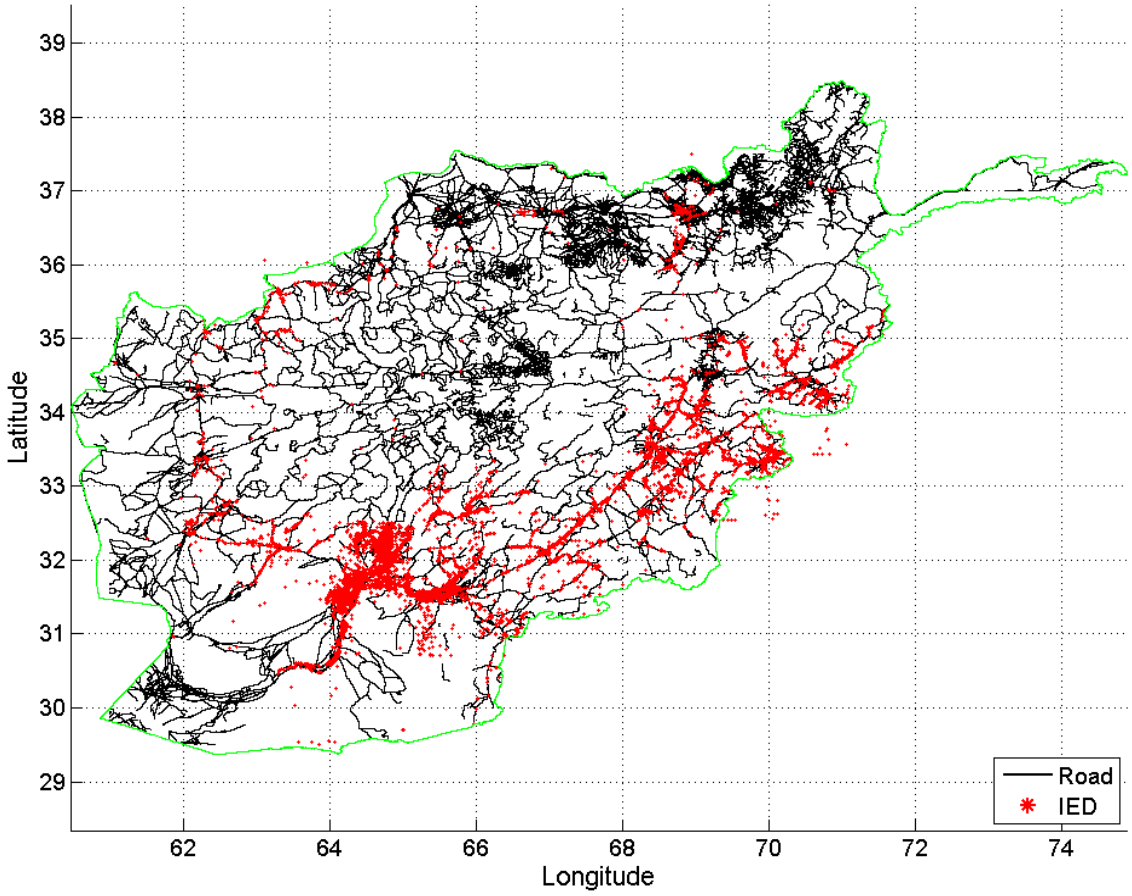



**Figure 1. IED attacks in Afghanistan, February 2011 - August 2012.**

extremes. IED explosive content has included fertilizer, rapidly oxidizing propellants, repurposed commercial explosives, gasoline, and artillery shells. IEDs are typically deployed in unconventional configurations that attempt to maximize concealment and lethality.

Roadside bombs are probably the most common use of IEDs in modern conflict. This implementation of the IED is triggered in several ways. Victim-detonated IEDs explode when the target actuates the trigger in some way, perhaps by stepping on it. Human-borne and vehicle-borne IEDs are triggered by the attacker accompanying the

device. Command-detonated IEDs, one of the more lethal configurations, are triggered by an attacker from a distance, using a command wire or radio signal to explode the device.

In Afghanistan, IEDs are one of the most lethal weapons, accounting for an estimated 1,337 deaths by the end of 2012 [2]. Figure 1 shows the distribution of IED attacks in Afghanistan from February, 2011 to August, 2012.

### *Direct Fire Attacks*

Ranging from a lone sniper concealed on a mountainside or in a building to a force-on-force battle, direct fire attacks are probably the most common conflict event. These attacks occur when the attackers fire weapons directly at the target. Direct fire attacks in asymmetric conflict are typically performed using rifles and pistols but may include any type of weapon that can be pointed directly at the target. Notably, artillery and mortars are not direct fire weapons. Direct fire attack configurations are limited only by the imagination and weaponry of the attackers. However, all direct fire attacks require that the shooter be able to see the target.

In Afghanistan, there are two common forms of direct fire attack. The most common form is probably engagement at a distance, where a small group of attackers engage a military patrol from a distance, attempting to kill or wound their targets through the use of high accuracy marksmanship. While this attack seldom yields high numbers of casualties, it slows the movement of the target. Also, this type of attack allows a small number of attackers to impede the efforts are a much larger target force. Another common form of direct fire is the ambush, where a group of attackers attempt to

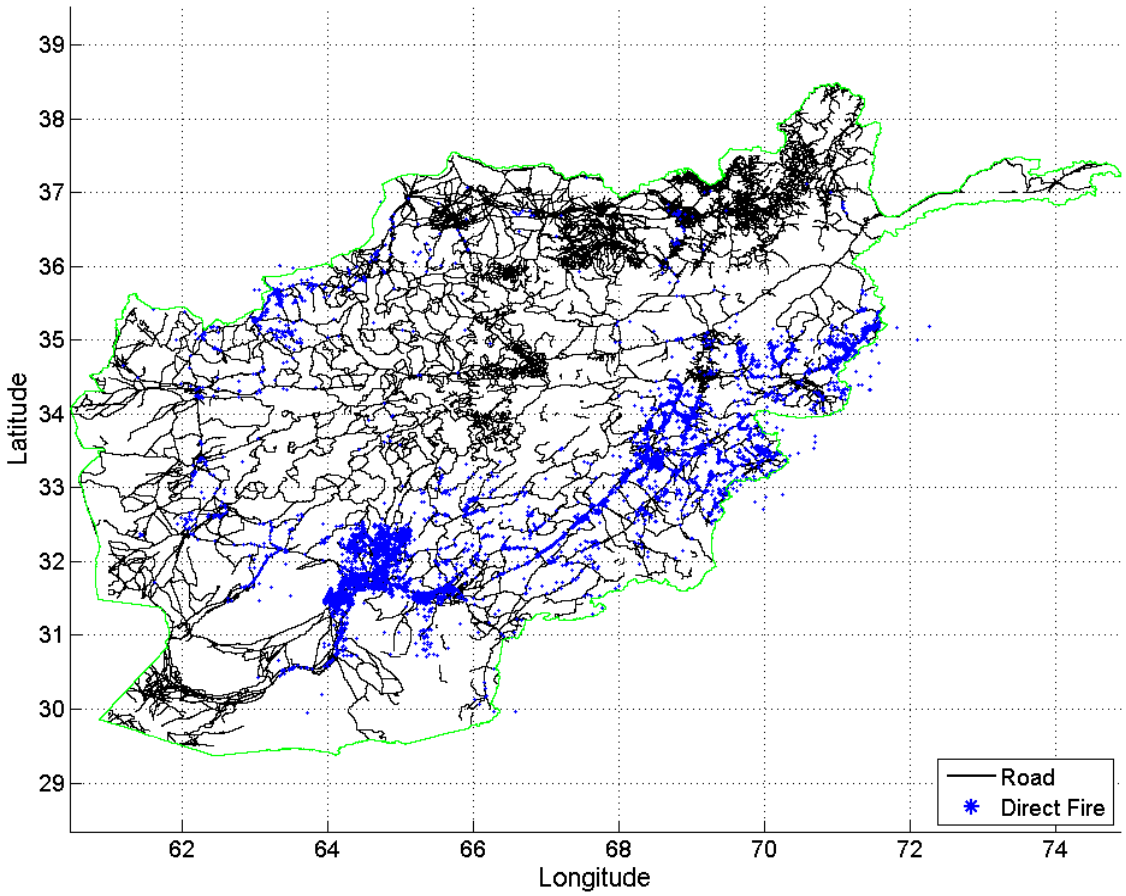



**Figure 2. Direct fire attacks in Afghanistan, February 2011 - August 2012.**

engage the target by launching an attack from concealment. In this case, the attackers will try to configure the attack in a way that guarantees numeric superiority and high rates of fire, at least during initial execution. Terrain support for this type of attack is important because attackers typically rely on concealment and the element of surprise to achieve success. A properly executed ambush often results in a large number of casualties due to the high rates of fire and the close proximity of the attackers to the target. Figure 2 shows the distribution of direct fire attacks in Afghanistan from February, 2011 to August, 2012.

## Organization of the Dissertation

After a discussion of Related Work in Chapter II, this study starts in Chapter III with the elaboration of a model that describes the event site, the surrounding terrain, and relationships between event actors, the event site, and the surrounding terrain. The essence of the proposed model is a combination of conventional terrain analysis and human use of terrain for conflict events.

Chapter IV describes the algorithms used to collect features that populate the model. Ranging from conventional terrain analytics to visibility-constrained features extracted from cumulative viewsheds, a subset of representative features are mathematically described and statistically analyzed. Our purpose is to demonstrate a cross-section of features and describe their relevance to the model. (For conciseness, only a subset of features is included in the body of the dissertation. All other collected features are presented in Appendix B.)

Predictive analysis of conflict events is examined in Chapter V. Classification of potential event sites is performed using discriminant analysis, support vector machines, and nearest neighbor algorithms. The topics of data normalization, spatial and temporal constraints, dimensionality reduction, feature selection, and classification algorithm parameter selection are also addressed.

Figure 3 offers an overview and roadmap of the dissertation.

The dissertation concludes with Conclusion, References and several supporting Appendices.

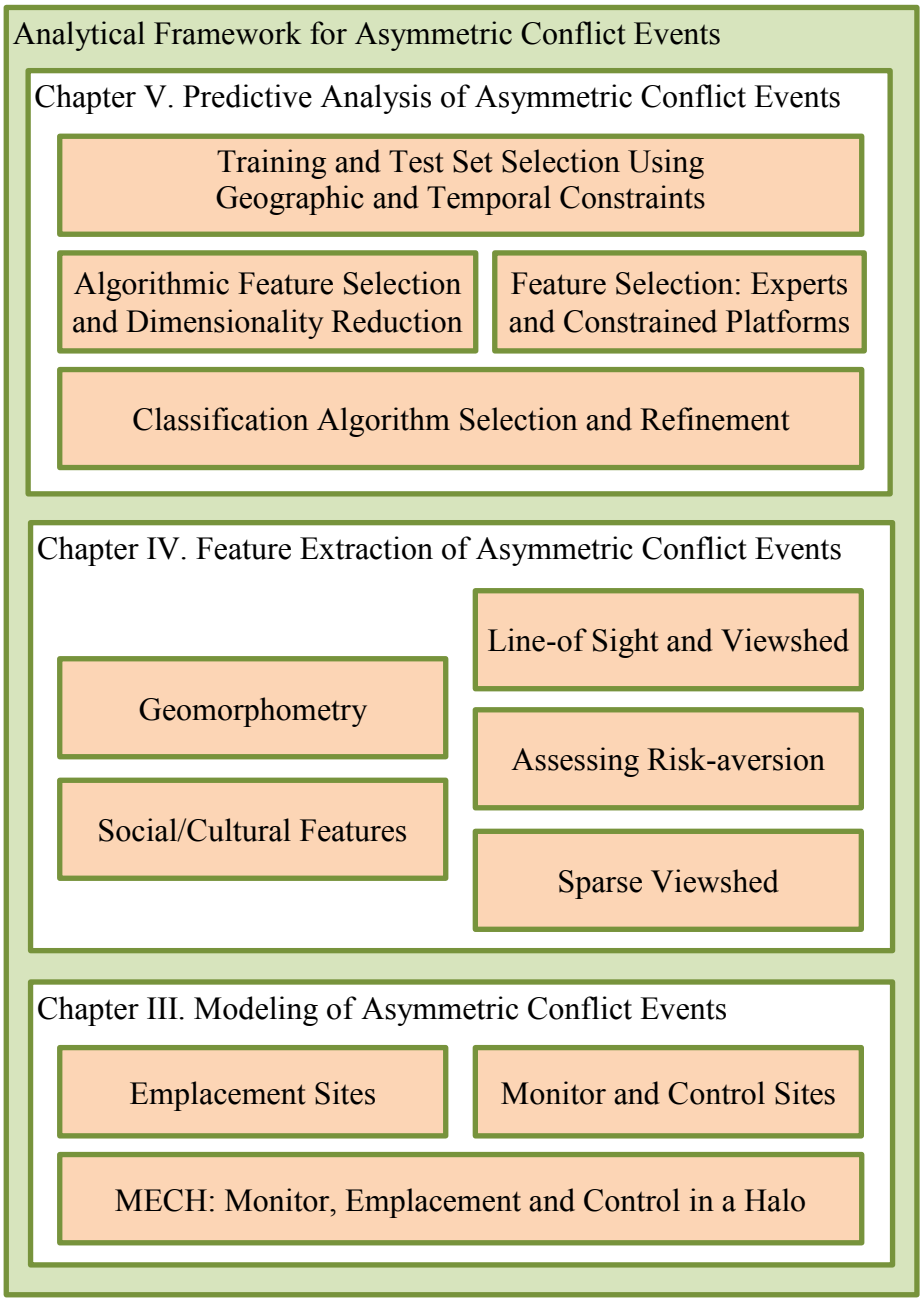



**Figure 3. Analytical framework for asymmetric conflict events.**

# CHAPTER II

# RELATED WORK

## Overview

Predictive analysis of conflict events incorporates elements from diverse domains. Related work falls into three categories: general conflict knowledge, feature extraction, and predictive analytics. General conflict knowledge is a catch-all category that includes information about friendly and enemy tactics, military and law enforcement processes and procedures, social and cultural behaviors of civilians in a conflict zone, and other topics useful in describing the expected actions, techniques, tactics and procedures of combatants and non-combatants. Feature extraction includes all efforts to add interpretation or information to raw measurements. Ranging from simple measurements and transforms to complex model-based assessments, features are an underlying and necessary component of every analytic effort supporting prediction. Finally, predictive analytics encompasses efforts made to find and mathematically describe patterns in historical data for the purpose of making predictions about the future (or unknown) data. These patterns may occur naturally, i.e. land contouring caused by running water, or may be the result of human action, i.e. positioning of an overwatch site to maximize field of view. In this research, mined patterns are generalized to models and used to classify the utility of locations for future attacks. The following sections offer a discussion of relevant readings and provide insight into published research related to these topics.

## General Knowledge

The generic term ‘conflict’ covers a wide range of activities commonly grouped into warfare, terrorism and criminal activity. Asymmetric warfare is the most prevalent source of conflict events in the modern world. Perry and Gordon provide an intelligence-oriented view in [3] focused on recent conflicts in Afghanistan and Iraq. They note the importance of understanding enemy intent and highlight substantive differences between conventional and asymmetric warfare, particularly with regard to localization of conflict tactics and adaptability of combatants. Townsley [4] provides a broader view of insurgency and supports with analysis the intuition that insurgent actions are based on rational choice and impacted by scarcity. Gutfraind takes a larger and more abstract view of terrorism in [5]. His terrorist network models offer insight into structures and vulnerabilities. Memon, et al. [6] contains a collection of peer-reviewed papers offering various mechanisms for modeling and analyzing terrorism. On a more realistic note, Lynn extracts common patterns from successful and unsuccessful insurgencies [7] based on the ideas of quantitative and qualitative violence. Of particular interest is the concept of qualitative violence where tactics are selected for maximum psychological effect instead of actual casualties or damage.

The problem of improvised explosive devices (IED) is pervasive on the modern battlefield. Testimony to the U.S. House Armed Services Committee [8] offers some insight in the nature of AW and its impact on U.S. forces with a particular focus on IEDs and related asymmetric warfare tactics. The focus on IEDs is advanced by Barker in [9] who finds that prevalence and effectiveness of IED events grew significantly between

2002 and 2009. His analysis covers the tactics and outcomes of two IED campaigns conducted by two different groups in Afghanistan and Pakistan and offered province-level event detail. Barker also notes the increasing speed of adaptation and information sharing on a global scale between insurgent and terrorist groups. Magnuson discusses the adaptability of the IED makers in [10]. In [11], Hanson examines the economic decisions that underlie the use of IEDs as a tactic and analyzes the impact of countermeasures. Notably, Hanson shows that countermeasures do not reduce the total number of IED attacks. IED effects and countermeasures are discussed at a high level in [12]. On a more tactical level, the IED Trigger Recognition guide [13] offers details about detecting and recognizing these devices for soldiers on the ground in Afghanistan.

Other types of conflict events are encountered in more mundane situations. Marijuana cultivation in national parks in the United States of America is an example of this type of conflict. For Yosemite National Park, the principle conflict is illicit marijuana cultivation. In [14], [15], Mallery examines the tactics, motivations and capabilities of drug trafficking organizations (DTO) that cultivate marijuana in national parks in California. Many of the choices made by DTOs with regards to marijuana cultivation site selection, infiltration and exfiltration routes, and site overwatch are similar to the choices made by attackers in asymmetric conflicts. For routine criminal activities that involve planning and site selection, like home burglaries, Risk Terrain Modeling [16] approaches predictive analysis with the concepts of opportunity theory, co-located events, and assessment of the social elements that make the crime possible.

An understanding of the tactics employed in asymmetric warfare is important. The U.S. Army Ranger Handbook [17] offers a concise compendium of small unit tactics and field operations. The U.S. military doctrine addressing counterinsurgency operations at higher levels is captured in [18].

## Feature Extraction

The selection of a site for a conflict event is the result of a complex, conditional decision-making process that includes terrain analysis, availability of necessary supporting infrastructure, opponent predictability and constraints imposed by event tactics. The extraction of these features from available or collected data is critically important to successful predictive analysis. Janlov, et al., discusses mechanisms to build situational awareness from these features in [19],

Geomorphometry is the science of quantitative land-surface analysis [20]. An interdisciplinary field, geomorphometry uses a variety of statistical and mathematical techniques, as well as algorithms drawn from image processing, to quantify characteristics of terrain and its underlying geometric structures. This analysis may take places at scales ranging from very small (on the order of sand grains) to very large (on the order of mountain ranges). Geomorphometric features are widely used in a variety of applications. One of the most common is terrain classification through the use of a geometric pattern, defined by Pike as “a set of measurements that describes topographic form well enough to distinguish geomorphically different landscapes” in [21], where he described as a set of landsurface geomorphometric parameters composed of 17 features derived primarily from topographic digital elevation models (DEM). Iwahashi and Pike

(IP) offer a simpler pattern in [22] that uses only three features. Stepinski and Bagaria extend IP's scheme in [23] by examining the three features over various window sizes. Other similar efforts include [24]–[28].

A key requirement of many offensive and defensive conflict tactics is the need for intervisibility, or line-of-sight (LOS), between the attacker and the target. Richbourg and Olson describe a variety of tactics that require LOS in [29]. Remotely triggered IEDs, ambushes, sniping, and even indirect fire all require the attacker to have an ability to observe the target at a distance. With this requirement, LOS becomes a useful analytical filter. However, computation of LOS to all points within an area, also known as viewshed, can be computationally expensive. Nagy [30] provides a basic description of viewshed and its utility. Floriani, et al, [31][32] offer an overview of intervisibility and a survey of intervisibility algorithms and problems. The Line of Sight Compendium [33], assembled by the U.S. Army Corps of Engineers, details 19 different LOS algorithms in use in various military simulations. Franklin examines optimal methods for large-scale viewshed analysis in [34] while Stiles [35] examines problems associated with the use and interpretation of viewsheds in tactical scenarios. Increased efficiency by dividing terrain in regions is the goal of Moet [36]. Geometric considerations, point interpolation and other sources of error are addressed by Guth in [37] while Raehtz examines the impact of error in DEM elevation estimates in [38]. Guth also addresses the impact of viewshed algorithm optimization choices in [39]. Izraelevitz offers a potentially lossy but fast viewshed algorithm in [40]. In [41], Washtell proposes, with little elaboration, the use of openness and geomorphometry to assess and classify

locations for certain purposes. Most LOS and viewshed algorithms, including those already mentioned, use the natural terrain surface as a base of calculations. However, in urban areas, this ignores the impact on LOS of buildings. Rød, in his exploration of visibility dominance in [42], offers ideas on incorporating artificial structures in viewshed computations.

Building on viewshed, and taking into consideration likely conflict tactics, a number of features may be constructed. Hengl, et al. [43] offer an excellent introduction and overview of standard geomorphometric measures. Terrain classification is addressed by Pike in [21], Iwahashi and Pike in [22], and by Stepinski and Bagaria in [23]. Openness, as described by Yokoyama in [44], incorporates terrain classification and elevation profile analysis and offers an interpretation of the three dimensional space visible from a given location. Guth uses openness to predict viewshed in [45]. In urban areas, the architectural concept of isovist may be more appropriate [46][47]. Another approach to mathematically defining terrain combines several features into a single terrain complexity metric. One example of a terrain complexity metric is proposed by Huaxing in [24]. Cumulative viewsheds are used to describe the overall exposure to view of a site from multiple locations. Wheatley applies this analysis to understand the siting of barrows [48] and Lake, et al. to Mesolithic village sites [49].

Other mechanisms exist that can be used to create features. Curtin proposes a shift from geographic coordinates to linear referencing in [50]. His approach reduces analysis by one dimension but fails to consider how sites away from indexed roads may contribute to events. In [51], aggregation techniques are used by Alt to develop binary,

count and duration models. Holte [52] notes that simple classification rules can be used to perform surprising complex analysis, especially on larger datasets.

The human mind has an innate ability to recognize subtle patterns and out-of-place elements. The sixth sense of the pointman who detects an ambush is an iconic Hollywood theme that is frequently reported in real life. In the movie "Platoon" [53], the pointman Chris failed to notice some off-color vegetation hiding an enemy bunker. Staff Sergeant Barnes intervenes and halts the patrol. From the movie script,

> *"Chris now starts to see things he didn't see. Right in front of his nose - there is a trench from this bunker to another and another. There is now in his view a complex of bunkers and thatched hootches and lean-tos all blending into the forest ..."*

In this scene, Staff Sergeant Barnes halted a patrol based on a feeling of danger or sixth sense. After the halt, Chris begins to consciously identify the threat indicators subconsciously detected by Staff Sergeant Barnes. Similarly, in [54], a British soldier named Kenny Meighan discusses his abilities to detect subtle inconsistencies, describing his success as a result of "… subconscious instincts that start to read what's going on …".

The placement of conflict events and the detection of these emplacements are probably strongly influenced by this sixth sense, intuition or instinct. It is unlikely that attackers conduct exhaustive map reconnaissance to select the perfect location for an attack. Instead, they identify a general area or stretch of road that provides necessary support for the attack and select the actual location based on intuition or its "feel".

Similarly, the pointman does not exhaustively consider every clue within eyesight. Instead, he instinctively recognizes terrain that is useful to attackers and unconsciously integrates indicators that might signal an attack. In [55], Gladwell discusses this intuition, describing it as "thin-slicing". Thin-slicing allows humans to form a snap judgment or immediate, unconscious reaction to a situation without consciously understanding or recognizing all of the inputs. Capturing the inputs to this type of instinctual decision is impossible, even for the person making it. However, it may be possible to summarize terrain by incorporating and transforming measurements based on sightlines and intervisibility to quantify feelings of openness and exposure, building on the insights of Yokoyama [44] and Guth [45].

## Predictive Analysis

Predictive analysis extracts information from data and uses it to predict future events. In the analysis of asymmetric conflict in this research, the information extracted reflects the tactics chosen by the attackers as revealed by their use of terrain and proximity to necessary support like populated areas. Based on features extracted from historical events, the development and use of predictive models covers a wide range of topics including normalization, dimensionality reduction, feature selection, and statistical learning.

### *Models*

Detection of patterns in battlefield data is a primary consideration in conflict event analysis. Patterns of interest are primarily caused by humans, both combatants and non-combatants. The National Research Council writes an overview of this topic in [56].

The book covers all aspects of modeling and simulating military operations including military and civilian organizations and individuals, cultural and social frameworks, and operations from individual movements to pitched battles between nation states. Deitchman [57] provides one of the seminal military papers in this field. His work describes the application of the Lanchester models of combat in common conflict scenarios covering both conventional and unconventional warfare. Focused only on personnel strengths at macro levels, the models are surprising effective at predicting the outcome of historical battles. In [58], Schaffer tailors the Lanchester models to address the characteristics of insurgencies where small insurgent units combat larger, more conventional forces. Although still focused primarily on attrition, Schaffer also addresses the impact of special tactics, weapons and morale. He also extends the model to specific types of engagement including the skirmish, ambush and siege. In [59], Kress examines the impact of intelligence on conflict outcomes. More recently, COIN 2.0 [60], described by Henscheid, captures the U.S. Department of Defense's state-of-the-art computational model to support military planning in counterinsurgency scenarios at all levels from strategic to tactical. COIN 2.0's breakdown of insurgent roles is useful in understanding how terrain might be used to support or conceal an attack. Bornstein, et al. further refine insurgent roles and specific indicators of adversarial intent in [61]. SCARE-S2 is used to find high-value targets in Afghanistan [62] using a simple set of event descriptors.

Combatants may reveal their intent and preferences through detectable choices they make. Similarly, combatant choices may be reflected in non-combatant activities.

An example of this is a classic signature of warfare: the abandoned village. Non-combatants flee areas where combatants are preparing for conflict. Johnson offers an overview of similar human-oriented patterns in [63]. This paper is updated and expanded in lecture notes prepared for a Fall 2011 seminar [64] where he offers a review of empirical patterns seen across a range of conflicts, both criminal and military. Modeling of more mundane and mechanical tasks is addressed by Okuda in [65]. In this paper he describes the sliding of a lever by an individual as a series of discrete tasks followed by a decision. Portland models the more complex system of an automobile and driver in [66]. Using a nested model approach, he ties together individual movements and choices for the purpose of predicting future driver behavior. In [67], Barrett models the infrastructure and users of the infrastructure in a medium-sized North American town and uses it to predict the spread of contagion, routing performance in a vehicle-based ad hoc network, and the impact of laws and new technologies on the electricity market. His models incorporate a high degree of complexity for a very large number of agents but are very computationally expensive to set up and run. Barrett's approach is expanded in [68] where he models the complex systems arising from the interaction of physical, biological, technological, informational, and human/societal components. This model forms the core of a simulator called Simfrastructure that can handle very large scale, detailed simulations.

In [69], Wilson et al. examine various analytic frameworks that may be used to analyze terrorism. Although narrowly focused on domestic terror and strongly slanted towards spread of contagion, this collection of papers examines scenarios beyond

today's battlefield. In [70], Sandler examines the impact of proactive counter-measures as part of a game-theoretic model. He also discusses the importance of information in asymmetric warfare but focuses on interactions between terrorists groups, their supporters and nation-states. Bohorquez [71] finds that event size and timing in modern insurgent conflicts follow a power law distribution,. His empirical analysis reveals that common drivers may underpin a broad swath of modern conflict. An empirically-derived model of IED emplacement is developed in [72] in order to support allocation of IED clearance assets. Although the model is limited to timing information and simple measures of success or failure, it does incorporate historical data that reflect the attackers' choices. Dixon, et al. introduce an agent-based coalescence-fragmentation model where individual agents can be people or resources [73]. This approach is applied to several variants of modern conflict. Conflict events are not the result of simple processes. Fabian [74] presents the problem as a two-stage recourse model that considers both feasibility and optimality simultaneously.

Modeling paradigms from related fields provide some insight. Jakobson offers a an overview of situation management in [75] and reasoning about complex, dynamic situations in [76].

Risk aversion is an important factor to consider in human conflict. Nikolova [77] and Swamy [78] examine the impact of risk aversion using combinatorial and stochastic optimization. Martinez [79] and Subrahmanian [80] focus on describing and mining human behavior using the Stochastic Opponent Modeling Agent (SOMA) . SOMA

offers a rules-based approach to predictive modeling of human behavior that has the ability to incorporate risk aversion.

### *Statistical Learning*

The predictive analysis in this research is founded on analysis of historical events to discover patterns. Statistical learning provides a foundation for many pattern recognition efforts. Hastie, et al. [81] offer an excellent introduction and elaboration of this field. In [82], Abu-Mostafa, et al., offer an overview of learning techniques and expose many common errors. Although simplistic in its approach, their book highlights subtle problems that lead to erroneous outcomes. Jain offers a short but comprehensive review of statistical pattern recognition in [83]. There are many pattern analysis mechanisms attempt to exploit spatial and temporal characteristics of the data. In [84] and [85], Ng offers a classic spatial data mining algorithm, CLARANS. Compieta takes a divide-and-conquer approach with spatio-temporal data in [86] while Kisilevich focuses on spatio-temporal clustering in [87]. Other clustering methods may also be useful. Globerson [88] formulates clustering as a convex optimization problem that attempts to collapse individual classes into single points. With neighborhood components analysis (NCA) [89], Goldberger uses a stochastic variant of k-nearest-neighbors to perform non-parametric clustering. NCA can struggle under high-dimensionality. Regularized NCA by Yang and Laaksonen [90] performs better at higher dimensionalities.

In this research, we compare and contrast the accuracy of classifiers built using support vector machines (SVM), k-nearest neighbors and discriminant analysis. SVMs

were originally proposed by Vapnik in [91], [92] and have been used in many applications including remote sensing [93], counterfeit banknote recognition [94], and to model interstate conflict [95]. Burges offers a well-known introduction to the use of SVM in [96]. LIBSVM, a library of SVM tools developed by Chang , provides a hands-on introduction to construction of SVM classifiers [97] that goes well with a practical guide offered by Hsu in [98]. A variety of kernels are used with SVM. One of the most common is the Gaussian radial basis function (RBF) which is used in terrain classification [99][100], land cover classification [101], traffic accident prediction based on date and weather [102], and spatial distribution of pollution [103]. All of these applications have significant geographic or spatial features that play an important role in classification. Parameterization of the RBF kernel is discussed in [104]–[106].

K-nearest neighbors (kNN) classifies samples based on a majority vote of the closest training examples [81] . A non-parametric machine learning algorithm, kNN is a simple but robust classifier with a bounded error rate [107]. kNN can be difficult when the number of features is large and when the number of training samples is small [108]. Random subspaces [109] by Ho manages high dimensionality by using randomly selected features to construct the classifier. Golberger, et al. use neighbourhood components analysis [89] to reduce dimensionality and learning the distance metric from the data itself.

In some cases, individual classifiers may be combined. Hastie, et al. [81] and Kittler, at al. [110] present broad views of different approaches to combining classifiers. Ting and Witten [111] offer insight into key problems and compare the results of well-

known ensemble learners. In [112] Fan, et al. compare simple classifier combinations and stacked generalization.

The ability to collect and analyze data sets with many features can lead to the curse of dimensionality and negative impacts on pattern recognition and statistical learning techniques. Selecting or creating good features is an important step in preparing the data for use in a machine learning algorithm. Guyon offers an introduction to feature selection in [113]. One interesting assertion made is that high correlation between features is not always a problem. Even highly correlated variables may have great complementarity. She also notes the importance of feature normalization, especially to iterative learners that seek convergence. Stepwise feature selection is a method of reducing dimensionality [114]–[116] that iteratively adds individual features, typically based on some statistical measure like the F-test. Stepwise feature selection has known drawbacks [117] and performs best when the features are normalized [113]. Principle component analysis (PCA) is another method of reducing dimensionality through the construction of linearly uncorrelated variables from existing, possibly correlated features [81]. PCA is sensitive to the scaling of the original features.

Data normalization is an important component of predictive analysis. In [118], van der Berh, et al., offers an overview and comparison of various scaling and transformation techniques. He finds that autoscaling and range scaling performed best for his data sets, particularly when combined with PCA. Jain notes the impact of outliers on common normalization schemes used in biometric systems in [119]. Karvanen examines the importance of normalization with samples are collected at different

locations by different people potentially using different instruments [120]. The problem of differing collection standards is probably prevalent in the collection of the conflict event data used in this research. In conflict, tactics change over time. Osgasawara, et al. propose and adaptive normalization scheme for handling non-stationary time-series data in [121]. Their approach uses moving averages to determine minimum and maximum values that can be used for min-max scaling.

Although the ability to collect data is great, in many cases, there may be few historical events to learn from. The early days of a conflict, the first forays into new territory, and the emergence of new tactics are all cases where the sample size available for training is small. King and Alt examine the impact of identifying rare events in [122]. Huang, et al. detect co-location patterns using rare events in [123].

**Summary of Open Problems**

A number of open problems must be addressed or improved in order to make accurate predictions about future human activities and choices in asymmetric conflict. These open problems drive the direction of this research effort. First, it is necessary to analyze terrain to extract the features used by the humans as they make their choices. This analysis should result in features that summarize terrain the way humans do, taking into account terrain utility, visibility and intervisibility, and other required support. Summarization must occur at scales and resolutions that are appropriate to both the feature being measured and the decision and the decision being made. Measurements associated with risk aversion must be developed. These measurements must be used in new features that incorporate risk aversion as a modifier or constraint.

Second, features should be incorporated into a model that operates at various resolutions in order to capture the conflict event site (within meters), nearby supporting terrain (within scores of meters), and its embedding into the local neighborhood (within kilometers). This model must be flexible enough to describe unknown tactics occurring across a wide variety of terrain types.

Finally, predictive analysis built on the model must be adaptive, incorporating spatial and temporal constraints on subset selection, dynamic normalization and feature selection and dynamic parameter and hyper-parameter determination. Normalization parameters must be learned from available data and feature selection or transformation must take into account geographic and temporal effects on tactics. Hyper-parameters, like the order of $k$ for kNN and the shape parameter $\sigma$ for radial basis functions, must also be learned from the data. Since this type of predictive analysis is useful in real-time on the battlefield, there is a need for a lightweight version of the predictive analysis system that provides best-effort results using constrained computing devices.

# CHAPTER III

# MODELING OF ASYMMETRIC CONFLICT EVENTS

Central to this dissertation is the development of a system model that captures human decisions and the interactions between attackers and terrain in the siting and execution of a conflict event. The model describes the combination of terrain and tactics that make a conflict event possible, including characterization of useful terrain and the use of terrain by attackers.

In military terms, tactics cover all aspects of the employment of units in combat. This includes the movement and arrangement of the personnel and resources involved with respect to terrain and opposing forces [124]. In this research, tactic will be similarly defined for asymmetric conflict events. A successful tactic maximizes the probability of success by making employment decisions optimized for the local terrain, the capabilities of the attackers, the vulnerabilities of the target, and the goals of the conflict event. In other words, the siting and execution of an event at a specific geographic location is constrained by tactics suited to the local terrain, appropriate for the opposing force, and within the abilities of the attacker.

Each conflict event is unique. However, all conflict events are planned and executed by humans. Many of these humans share common training and similar experiences. These humans are likely to make similar decisions when faced with similar choices. The conflict environs also impose constraints. At a given time and place in a particular conflict, attacker access to conflict tools and weaponry is likely to be similar.

Terrain constrains tactic choice similarly across the conflict area. Target capabilities and vulnerabilities will also tend to constrain tactics, especially as countermeasures emerge for classes of conflict events. These shared elements enable predictive analysis of conflict events.

## Anatomy of a Conflict Event

The planning and execution of a conflict event involves a series of choices made over a period of time. These decisions are primarily concerned with selecting a location that supports execution of some particular tactic. The decisions address three conflict event elements. The Emplacement site is the place where the event occurs. For an IED, this is the location where the device is concealed. For a direct fire event, the Emplacement site is the center of the targeted force. The Monitor location is used for overwatch and early warning and will typically have good visibility of terrain along the approaches to the Emplacement site. The Control location is used to initiate execution of the conflict event and will typically have good visibility of the Emplacement site and adjacent terrain.

For this research, we believe that the planning and execution of a conflict event is accomplished in series of steps. First, the conflict event planner selects a particular class of event to execute, like IED or direct fire, and a general area that is likely to be well-suited for the event being planned. Factors involved in the selection of the area probably include the availability of targets, known or suspected availability of useful conflict event sites, and proximity to necessary support structures like population centers and communications networks.

At this point in the conflict event planning effort, a class of event has been selected. However, availability of specific supporting features in the general area may constrain the choice of tactics. Next, a planner travels to the general area and selects a specific site. Site selection starts with analysis of potential Emplacement locations. This analysis primarily addresses the utility of the terrain at the site for the type of event chosen and the general embedding of the site in the local terrain. Useful locations are further analyzed for availability of Control and Monitor locations. This utility of these locations is primarily a function of their intervisibility with terrain at and adjacent to the Emplacement site. Adequate Emplacement sites with adequate Control and Monitor locations are identified. One of these sites is selected and the Emplacement occurs. Notably, it seems unlikely that the choice of Emplacement site is optimal for a given area. We believe that planners select sites that meet all required criteria but do not have the ability to exhaustively analyze every possible combination of Emplacement sites and Monitor and Control locations across a general area in order to make an optimal choice.

Once a conflict event location has been selected, the Emplacement site is prepared and human actors are placed where needed. Actors at Monitor sites provide overwatch and early warning. The conflict event is initiated by the Control actor when a suitable target reaches the Emplacement site.

## The MECH Model

The MECH model is composed of conflict event features that represent Emplacement and Monitor/Control locations. These features capture the outcome of complex decisions made in the planning and execution of tactics. The features are

collected into tactical patterns at various resolutions: at the Emplacement site, immediately adjacent to the Emplacement site; and within the Halo, the annular area centered on the Emplacement site that can be used to perform Monitor and Control functions.

*Emplacement Modeling*

As previously described, the selection of a site and the execution of a conflict event is often the result of a carefully planned process. Emplacement at a site includes all activities required to select and prepare the location for the event. It also includes the relationship of the site to nearby and surrounding terrain. Let $R = \{r_1, r_2, \dots, r_n\}$ be a set of past conflict event locations. Let $X^E(r_x) = \{x_1, x_2, \dots, x_m,\}$ be a set of $m$ geomorphometric and other measurable features at the location $r_x \in R$. Then the tactical pattern $\tau_E$ of the Emplacement site $r_x$ is defined as the vector

$$\tau_E(r_x) = \left[c_j f_j\left(X_j^E(r_x)\right)\right] \qquad \forall\, j\ in\ X^E(r_x) \tag{1}$$

, where $c_j$ is a weight coefficient for feature $j$ and $f_j\left(X_j^E(r_x)\right)$ is the score of feature $j$ for location $r_x$. Feature $j$ is at or adjacent to the Emplacement site.

*Monitor/Control Modeling*

Actors play important roles in the execution of conflict events. For example, a carefully timed ambush only succeeds if the triggerman can observe his target's movements without being detected as an attacker before the attack is launched. Two roles common to many conflict events are the Monitor and the Control. The Monitor observes the target at a distance, provides overwatch, and reports to the Control. The Control observes the target and directs the execution of the conflict event. Note that in

some cases, a single actor may perform both Monitor and Control functions from one or more locations.

Let $H$ be an annulus, or Halo, with a variable inner boundary that may approach zero and a variable outer boundary that may approach the maximum limit of intervisibility with $r_x \in R$. This maximum limit may be the absolute physical limit of aided or unaided human eyesight. More typically, this maximum limit will be related to details of the conflict event task being performed. In this dissertation, we assume that an actor at a Monitor location must be able to distinguish between targets and non-targets and report target activities while observing from the outer bound of $H$ using unaided eyesight. We estimate this distance to be approximately 1500 meters and thus define its maximum value as 2500 meters for the case of Afghanistan. Define $X^H(r_x) = \{x_1, x, \dots, x_m\}$ be a set of $m$ Monitor and Control features measured over or extracted from terrain within $H$. Then the tactical pattern $\tau_H$ of the area surrounding $r_x$ encompassed by $H$ is defined as the vector

$$\tau_H(r_x) = \left[c_j f_j\left(X_j^H(r_x)\right)\right] \qquad \forall\, j\ in\ X^H(r_x) \tag{2}$$

, where $c_j$ is a weight coefficient for feature $j$ and $f_j\left(X_j^H(r_x)\right)$ is the score of feature $j$ for location $r_x$. Feature $j$ is measured over or extracted from terrain within $H$.

Once the Emplacement and Monitor/Control models are complete, it is necessary to model their interaction in order to accurately characterize their relationship in the execution of specific conflict events. The tactical pattern of the conflict event $T(r_x)$ , is the vector

$$T(r_x) = [\tau_E(r_x) \quad \tau_H(r_x)] \tag{3}$$

### *MECH and Tactics*

Attackers are faced with scarce resources and have a desire to optimize outcomes. Thus, an attacker with a specific goal, e.g. targeting opposing forces with an IED, will attempt to maximize success by making good choices. This does not mean that every choice is optimal. However, in the eyes of the attacker, the attack configuration for each specific conflict event will be good enough to succeed given the resources, training and time available to the attacker.

Attackers are assumed to have some level of training and experience. They are also assumed to be familiar with the area local to the attack, although the familiarity may be cursory or limited. Components of a successful attack, particularly Control and Monitor locations, may be reused. Likewise, successful attacks may be replicated on distant but similar terrain. Replication may also be a function of training, where successful tactics and adaptations are communicated to distant groups [125], [126].

The concepts of cursory familiarity and attack replication expose MECH's underlying assumptions about attacker methodology and abilities. MECH does not assume that attackers use detailed geographic maps or perform exhaustive analysis of local terrain prior to the placement of an attack. Instead, as described by Gladwell, they rely on experience and instinct. Attack emplacement is done at a place that 'feels right' or 'looks right'. This feeling is the result of both conscious and subconscious processing of the geometric structure of the local terrain, sight lines to prominent or useful terrain features, proximity to necessary logistical support, similarity to past successful conflict event sites, and similarity to a mental model of a 'good' site. A location with the right

'feel' is further assessed for critical attack support structures, like adequate concealment for an ambush or acceptable IED Control sites within range of the available command detonation wire. A successful attack confirms the 'feels right' analysis and solidifies the attacker's intuition. Thus, the general shape and configuration of an attack, a tactical pattern captured by Equation 3, may be mapped onto new locations (roughly replicated) in order to duplicate previous successes.

Pattern drift is a side effect of replication. New locations are never exactly the same as previous locations. Attack parameters must be shifted to make the old pattern fit the new location. When these adjustments are made and a successful attack occurs, the pattern grows or shifts. The result is a change in tactics over time.

Occasionally, a pattern will lose effectiveness. This may occur due to countermeasures, like new IED detection equipment, or due to a lack of critical attack components, like a particular type of IED detonator. When this happens, an abrupt shift in tactics may be seen.

# CHAPTER IV

# FEATURE EXTRACTION OF ASYMMETRIC CONFLICT EVENTS

A principle component of MECH-based analytics is extraction of features relevant to the siting and execution of a conflict event. These features are drawn from three general classes: visibility-related featured based on characteristics of the Emplacement site's viewshed; geomorphometric features based on characteristics of the land's surface; and social/cultural features related to proximity of human population centers.

Visibility-based analysis attempts to use human factors and limitations to constrain areas under consideration. For example, if an IED uses a command-detonated trigger, then it is likely that the Control site has direct line-of-sight (LOS) to the Emplacement site. Potential Control sites that cannot see the Emplacement site are probably less useful. Conversely, an ambush that relies on attacker concealment probably requires a relatively large area near the Emplacement site that is concealed from target view. A potential Emplacement site without nearby concealment is less useful in this scenario. Visibility-based analysis also attempts to summarize the impression of a location gained by a trained attacker during site assessment. Visible areas and inferred hidden areas are assessed and mentally summarized by the attacker at various scales related to the planned attack. Local viewshed and related features attempt to capture this assessment process.

Geomorphometry is the science of quantitative land-surface analysis [43]. For MECH, geomorphometric features are drawn from statistical analysis of the ASTER Global Digital Elevation Map (is a product of METI and NASA) [127]. Various morphologic features describing terrain structure and capturing terrain surface information are collected using elevation data with a resolution of 30 meters. The features are collected at a variety of windows sizes.

For both geomorphometric and visibility-related features, analysis at various window sizes is necessary. Window size determination is an open problem in geomorphometry [128]. Although several automated and semi-automated approaches have been advanced (variograms are frequently proposed [129]), the most common solution is to manually assign fixed window sizes large enough to contain the features and activities being analyzed [130]. For MECH, a total of six window sizes have been assigned, based on known or estimated attacker requirements. Table 1 lists and describes these window sizes.

For each of the collected features, we show a boxplot and a smoothed histogram in which three different Afghanistan data sets are compared: roads and two classes of events: improvised explosive devices (IED) and direct fire (DF). The IED and DF datasets are composed of conflict events that occurred in Afghanistan between early 2011 and mid-2012. See Appendix A.2 for more information. The roads set is composed of discrete points selected at 30 meter intervals from paved and improved roads across all of Afghanistan. See Appendix A.3 for more information.

**Table 1. Window sizes used for geomorphometric and visibility-based analysis**

| Radius (meters) | Rationale |
|---|---|
| 25 | The radius of the area considered immediately adjacent to the Emplacement site. Also, estimated to be the maximum range of a typical IED blast. |
| 100 | The radius of the area surrounding the Emplacement site that is useful for a near ambush or similar direct fire event. |
| 350 | The radius of the area surrounding the Emplacement site that is useful for a far ambush or similar direct fire event. Also, a typical limit of long rifle suppressive fire. |
| 500 | The radius of an area surrounding the Emplacement site that is useful for Control functions, like command wire detonation of an IED. (Estimated anecdotally.) |
| 1000 | The radius of an area surrounding the Emplacement site estimated to be most useful for Monitor functions. Also, a typical limit of crew-served weapon suppressive fire. (Estimated anecdotally.) |
| 2500 | The maximum sight line considered in this analysis. |

In the following analysis, boxplots and histograms are used to concisely describe the features. For the boxplots, the whiskers extend to the most extreme data points that are not considered outliers. Outliers are not displayed. The notches on either side of the median can be used to understand similarities between samples. Two samples are probably drawn from different populations (significantly different at $\alpha=0.05$) if their intervals (the width of the opening of the notch) do not overlap. In addition to

conventional boxplot information, there is an additional symbol in each boxplot (diamond, square, and circle) located at the mean of the data.

The histograms were generated using either 100 bins or bins at 30-meter intervals (to coincide with resolution of some of the data). Since the number of samples varies greatly between sets, e.g. ~3.4 million road points vs. ~13,000 IED events, the resulting bin magnitudes are represented as a percent of the total sample. Thus, the magnitude of a bin containing 41,000 road points would be 100*(41k/3.4M). Bins are represented by a point at [bin center, magnitude] and adjacent points are joined with a line segment to facilitate visual interpretation.

The following sections detail an illustrative subset of features used in the MECH model. For conciseness in the body of this dissertation, other features collected and used in MECH can be found in Appendix B.

**Visibility-related Features: Line-of-Sight and Viewshed**

Line-of-sight (LOS) describes the intervisibility between two points: if the points are visible to each other, they have LOS. Intervisibility is a common requirement for many conflict event activities. The Control actor intending to accurately trigger an IED needs intervisibility with the Emplacement site. The sniper needs intervisibility with the target in order to fire accurately. The addition of an LOS constraint to geomorphometric features allows the interpretation of those features in a new way. In some cases, LOS may be interrupted by nearby terrain. In these cases, activities that require LOS cannot occur beyond the interruption or break in intervisibility. Thus, terrain beyond the break can be excluded from analysis related to activities that require LOS. Multiple LOS may

be combined into a viewshed, which may be exhaustive or sparse. Features can be collected directly from both types of viewsheds. Viewsheds may also be used as a constraint or mask for the collection of other features. Finally, multiple viewsheds may be combined and features then collected from the result.

As presented below, this analysis offers a greedy assessment of viewshed. The impact of DEM error, surface irradiance, precipitation, dust, and vegetation are not assessed and would tend to degrade or change the estimate of visibility. Also, the viewsheds are estimated using elevation data at a resolution of 30 meters. The resolution is probably insufficient to capture some significant viewshed details.

*Line of Sight*

Denote as $r_x$ a location of interest and let $P = [p_0\, p_1 \ldots\, p_m]$ be a vector of $m$ points evenly distributed along a line extending from $r_x$ to a distant point $DEM_{m,n}$ such that $p_0 = r_x,\ p_m = DEM_{m,n}$ where *DEM* is a digital elevation map. Let $L$ denote the vector of elevations for $P$. These elevations may be adjusted for observer height, $h$..

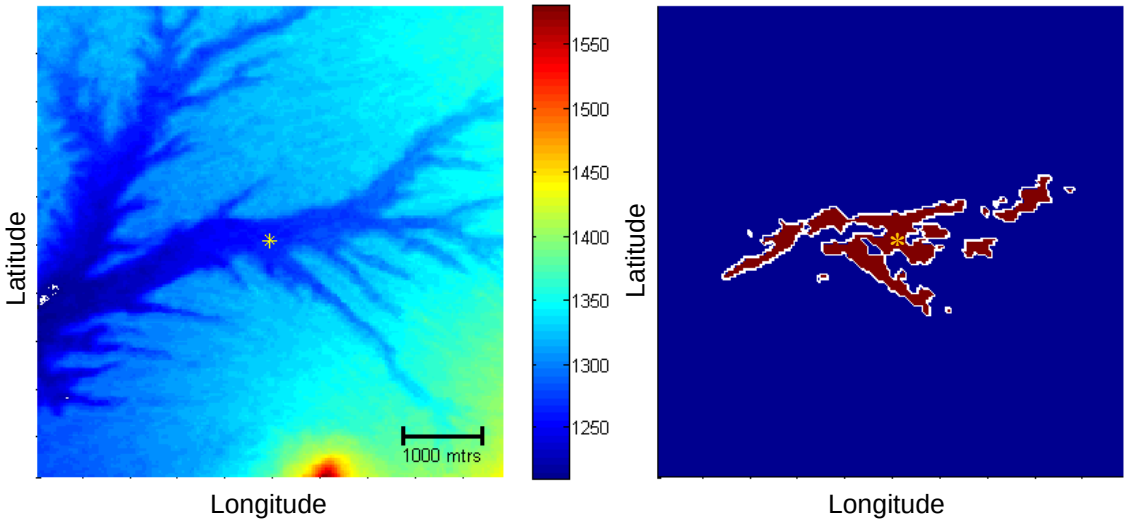
$$L = [l_0\, l_1\ \ldots\ l_m], \quad l_0 = elevation(r_x), \quad l_j = elevation(p_j)\ for\ j = 1, \ldots, m \quad (4)$$

Then, LOS between $r_x$ and all points in $P$ is the vector

$$LOS(L) = \left[slope(l_0 + h, l_j + h) \geq max\left(slope(l_0 + h, l_{1,\ldots,j-1})\right)\right] \quad (5)$$

$$for\ \ j = 1, \ldots, m$$

$LOS(L)$ is a vector of Boolean values that describes the intervisibility between $r_x$ and each of the points in $P$. This is similar to the approach adopted by Izraelevitz in [40].




(a) (b)

**Figure 4. The relationship between elevation and viewshed; a) An elevation map and b) its associated viewshed. A yellow asterisk denotes $r_x$.**

*Viewshed*

A viewshed describes the portion of a geographic area visible from a single point. Viewshed is calculated by determining LOS between a site and a set of surrounding points. Viewsheds are a useful way to see terrain through the eyes of an attacker. Hidden or revealed terrain and openness or exposure of a site are examples of information visible in a viewshed that is difficult to see in a conventional elevation map.

Figure 4.a offers an elevation map and its associated viewshed (Figure 4.b). The location of interest, $r_x$, is marked with a yellow asterisk in the center and the viewshed is calculated with respect to this location. In the viewshed, locations with LOS to $r_x$ are marked in red. Locations without LOS to $r_x$ are marked in blue. Locations marked in white are at the edge of intervisibility. Interestingly, it is difficult to gain an understanding of the viewshed from visual inspection of the elevation map alone.

Calculation of a viewshed provides an understanding of the intervisibility of $r_x$ with the surrounding terrain. This understanding contributes to remote assessment of the site regarding its vulnerability to certain types of attacks or its potential exposure to hostile observers. Any action directed against $r_x$ that requires intervisibility must originate or be triggered from a location within the $r_x$ viewshed.

There is a wide variety of viewshed algorithms. A cross section of these algorithms can be found in [40], [131], [132]. In this dissertation, a radial sweep algorithm similar to [131] is used to determine viewsheds. Two drivers inform this choice: optimization and feature extraction. Since radials can be processed independently or in batches, a high degree of parallelization may be achieved and the parallelization can be tailored to the number of threads or processes available. This allows more optimal use of available computing resources. This optimality is important because exhaustive viewshed determination is computationally expensive. The second driver is feature extraction. Subsets of radials can be used to extract features that summarize terrain geometry at various degrees of compression to succinctly capture visibility-related characteristics.

In order to determine viewshed using a radial sampling algorithm, denote the location of interest as $r_x$. Define radius *rad* to be the length of the maximum possible sightline of interest. Define $N_S$ as the number of radials that will be used to determine the viewshed. As a rule of thumb, for an exhaustive viewshed where LOS is determined between $r_x$ and every pixel within radius *rad*, $N_S$ should be the number of pixels on a circle of radius *rad*,

$$N_S = \lceil 2\pi * rad \rceil \quad (6)$$

Note that *rad* constrains the area considered to be part of the viewshed. There may be points at distances greater than *rad* that have LOS with $r_x$. Also note that for some features, $N_S$ may be set to sample radials much more sparsely. The impact of this choice will be explored for some sparse viewshed visibility-related features.

Let Θ be a vector of size $N_S$ consisting of angles, evenly spaced between $2\pi/N_S$ and $2\pi$, where $\Theta = [\theta_1, \theta_2, \ldots \theta_{N_S}]$, and $\theta_i = i * 360/N_S°$ . Then, $P_{\theta_i}$ is a vector of points distributed along a radial line extending outward from $r_x$ at angle $\theta_i$ to a distance of *rad* and the elevations of the points are denoted as $S = [L_1, L_2, \ldots, L_{N_S}]$ where $L_i$ is the elevation vector of $P_{\theta_i}$ and given by

$$L_i = [l_{i,0}\ l_{i,1}\ \ldots\ l_{i,m}], \qquad l_{i,0} = elevation(r_x), \qquad l_{i,j} = elevation(p_j) \qquad (7)$$

$$for\ j = 1, \ldots, m\ ,$$

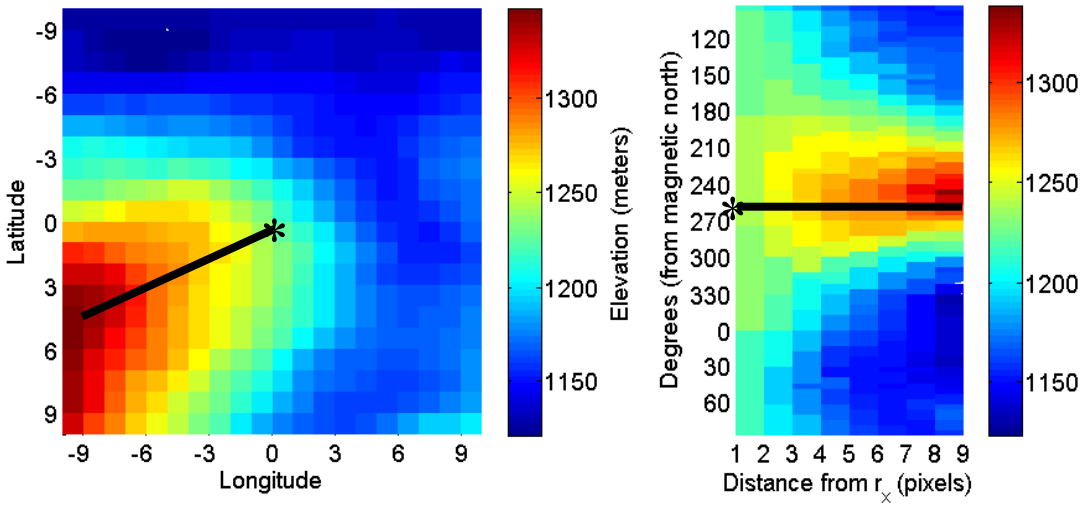
and $p_j$ is the coordinates of a point in $P_{\theta_i}$ defined by $\left(\left\lceil \frac{rad}{m} * j * \cos\theta_i \right\rceil, \left\lceil \frac{rad}{m} * j * \sin\theta_i \right\rceil\right)$. Then the sampling matrix $S$ of a circular area around $r_x$ is translated into matrix form, where the *i*th row of $S$ represents elevations for the points in $P_{\theta_i}$ and the *j*th column of $S$ represents the pixel distance from $r_x$ given by $\frac{rad}{m} * j$.

LOS along radial $P_{\theta_i}$ is described similar to Equation 5:

$$LOS(L_i) = \left[slope(l_{i,0} + h, l_{i,j} + h) \geq max\left(slope(l_{i,0} + h, l_{(i,1),\ldots,(i,j-1)})\right)\right] \qquad (8)$$

$for\ j = 1, \ldots, m$ , where a value of TRUE indicates that the point has LOS with $r_x$.

Viewshed is determined by finding $LOS(L_i)$ for all desired radials




(a) (b)

**Figure 5. Conventional and radial elevation maps; a) A conventional elevation map; b) the same elevation map, presented in radial format. The black line denotes the same radial in both plots.**

$$VS(r_x) = LOS(L_i) \quad for\ i = 1, \dots, N_S \tag{9}$$

Figure 5.a displays a conventional elevation map. The x-axis and y-axis represent longitude and latitude, respectively, and the colors indicate elevation. The asterisk at coordinate [0,0] represents $r_x$. Figure 5.b is the same elevation map displayed in radial form. Each row contains the elevation vector $L_i$ for a single $P_{\theta_i}$. Each column contains the $N_S$ elevations of points on the circumference of a circle centered on $r_x$. In Figure 5.a, the black line starting at $r_x$ and extending outward along an azimuth of approximately 250° magnetic represents $P_{250}$. The black line in Figure 5.b is also $P_{250}$.

*Viewshed Feature: Visibility Index*

Visibility index describes the total amount of terrain within a specified radius or window that is visible from the center of the viewshed. Visibility index provides a scalar

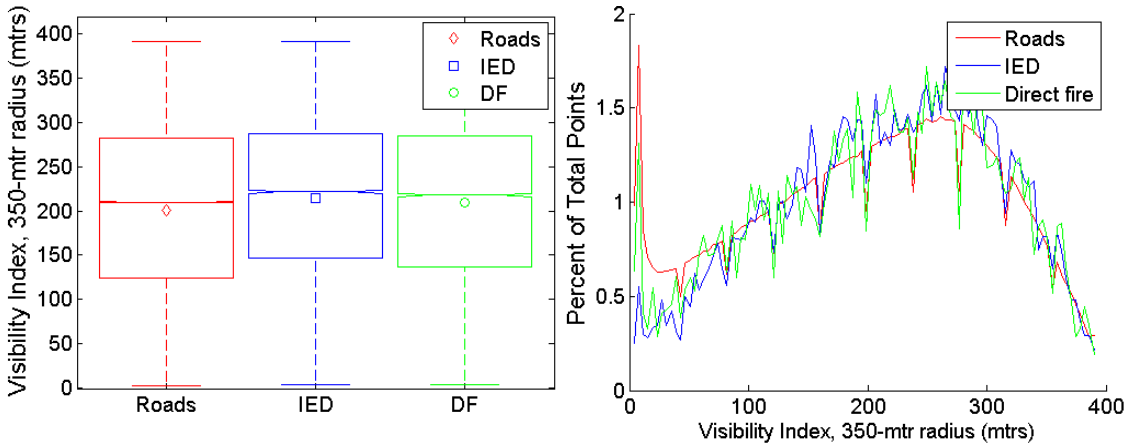



**Figure 6. Visibility index at a radius of 350 meters.**

assessment of the total intervisibility of $r_x$ and the surrounding terrain and gives an indication of the degree of exposure or concealment of the site. Given a conventional exhaustive viewshed $VS(r_x)$, as depicted in Figure 4.b,and a radius $w$,

$$VisIndex_w(r_x) = \sum_{i,j} VS(r_x)_{i,j} \mid distance\left(r_x, VS(r_x)_{i,j}\right) < w \qquad (10)$$

In Figure 6, visibility index is calculated from an exhaustive viewshed over a radius of 350 meters. Notably, there is a significant degree of overlap between roads and all classes of events. Visibility indices for other window sizes are in Appendix B.1.

*Viewshed Feature: Discrete Shape Complexity Index*

The Discrete Shape Complexity Index ($SCI_D$) describes the general complexity of a viewshed by capturing how dispersed it is. $SCI_D$ is derived as a perimeter-to-boundary ratio. The perimeter $P$ is the actual count of pixels with LOS that are adjacent to pixels without LOS. In terms of the exhaustive viewshed, this means that at least one of the eight adjacent pixels has no LOS with $r_x$. The boundary is a circumference of the smallest circle whose area equals the total count of pixels with LOS to $r_x$.

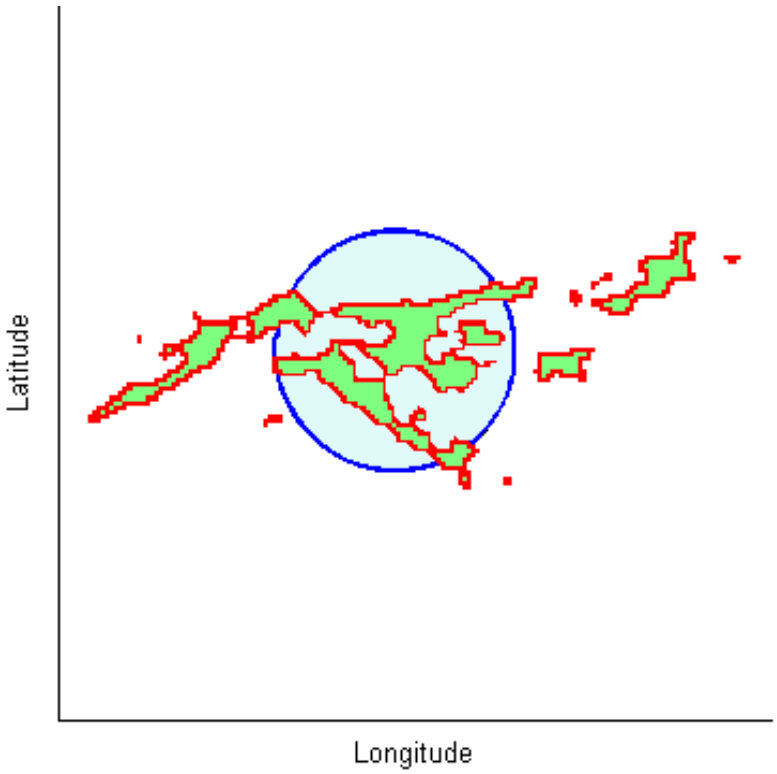



**Figure 7. Viewshed and discrete shape complexity index.**

In Figure 7, points within the viewshed are colored green and red. The red pixels denote the perimeter of the actual viewshed. Each red pixel is adjacent to at least one pixel that does not have intervisibility with $r_x$. The dark blue circle encloses an area equal in size to the total area represented by the red and green pixels.

$SCI_D$ is found as

$$SCI_D(r_x) = \frac{P}{2 * \pi * r}, \qquad r = \sqrt{\frac{\sum_{i,j} VS(r_x)_{i,j}}{\pi}} \; \forall \; i,j \; in \; VS(r_x) \tag{11}$$

$$P = \sum_{i,j} \left( VS(r_x)_{i,j} \equiv 1 \quad \& \quad \left( \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} VS(r_x)_{m,n} \right) < 9 \right) \; \forall \; i,j \; in \; VS(r_x) \tag{12}$$

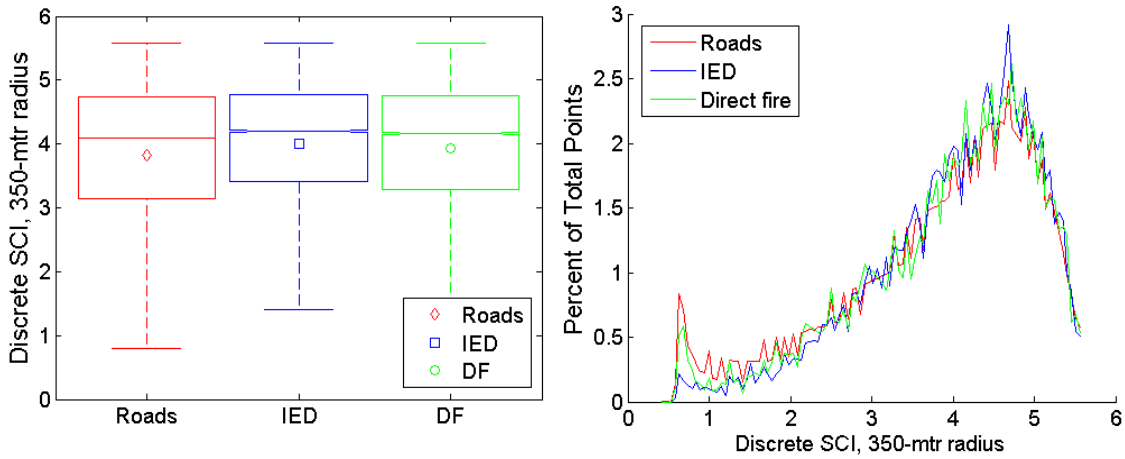



**Figure 8. Discrete Shape Complexity Index at a radius of 350 meters.**

Note that in Equation 12, if $VS(r_x)_{i,j}$ has a value of 1 (has LOS with $r_x$) AND at least one surrounding pixel does not (no LOS with $r_x$), then the result is a 1.

Figure 8 shows $SCI_D$ calculated across a window with a radius of 350 meters. The distribution of road points and events appears to be similar. $SCI_D$ for other window sizes can be found in Appendix B.2.

### Risk Aversion and Escape Adjacency

Many types of conflict events rely on concealment, camouflage, and the element of surprise. The success of IED attacks frequently depends on the ability of the actors to remain hidden until the target is correctly positioned and the attack is launched. Similarly, direct fire attacks may last longer or be more effective if the shooters can fire from a protected location. Success for both types of events requires the actors to avoid risk as much as possible before the attack. Thus, an important element of conflict event site selection is the identification of locations around the conflict event that offer cover and concealment to a risk averse attacker. In this case, the concept of *escape adjacency*

may provide insight into the tolerance of risk by an actor. A location with escape adjacency has LOS with $r_x$ but is adjacent to a location without LOS to $r_x$. These locations lie on the perimeter of the viewshed and are marked in red in Figure 7. When situated at an escape adjacent location, an actor can move quickly from visible to hidden with regard to $r_x$. Similarly, an actor can position himself at the edge of intervisibility in an effort to jointly maximize visibility of the target and concealment from the target. A risk averse attacker might prefer locations with escape adjacency.

Let $VS(r_x)$ be a viewshed, organized as an $n \times n$ binary matrix with $r_x$ at its center, as depicted in Figure 4.b. Then the escape adjacency matrix $EA(r_x)$ defined over $VS(r_x)$ is

$$EA(r_x) = \left[ \quad \left(VS(r_x)_{i,j} \equiv 1\right) \quad \& \quad \left( \left( \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} VS(r_x)_{m,n} \right) < 9 \right) \quad \right] \qquad (13)$$

$$\forall\, i,j \; in \; VS(r_x)$$

In $EA(r_x)$, the resulting $n \times n$ matrix, escape adjacent pixels are those that (1) have LOS with $r_x$; and (2) have at least one neighboring that does not have LOS with $r_x$. Condition (1) is satisfied when $VS(r_x)_{i,j}$ has a value of 1, indicating that the location has intervisibility with $r_x$. Condition (2) is satisfied when the sum of $VS(r_x)_{i,j}$ and all adjacent pixels is less than nine, indicating that at least one of the neighbors does not have intervisibility with $r_x$. These are the same conditions used in Equation 12.

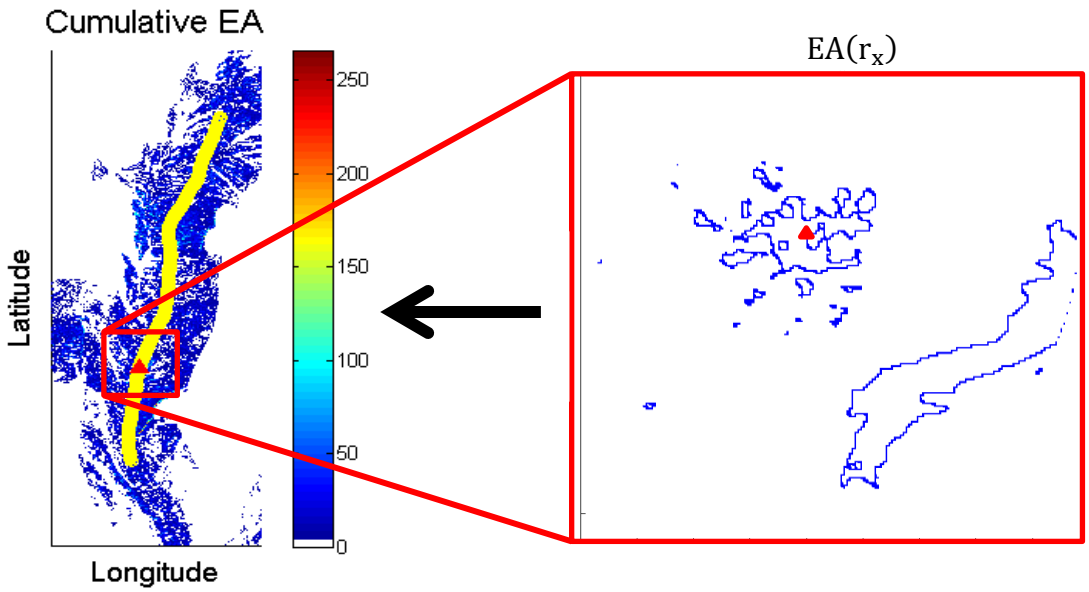



**Figure 9. Cumulative escape adjacency (CEA); a) CEA along route *R*; b) Escape adjacency for a single $r_x \in R$.**

*Cumulative Escape Adjacency*

Once escape adjacent locations have been defined for all $r_x \in R$, cumulative escape adjacency (CEA) can be calculated. CEA is determined by overlaying onto a single map the escape adjacent locations for all $r_x$ along a road or route.

Let $R = [r_1, r_2, \ldots, r_n]$ be a route or roadway of interest divided into *n* discrete points at a constant interval. Let $CEA(R)$ be a geographically localized two-dimensional zero matrix sufficiently large to encompass all terrain within some specified distance of every point in *R*. Then the cumulative escape adjacency map for the route *R*, $CEA(R)$, is the summation of the individual $EA(r_x)$ maps for each $r_x \in R$.

$$CEA(R) = CEA(R) + EA(r_x) \qquad \forall\, r_x \in R \tag{14}$$

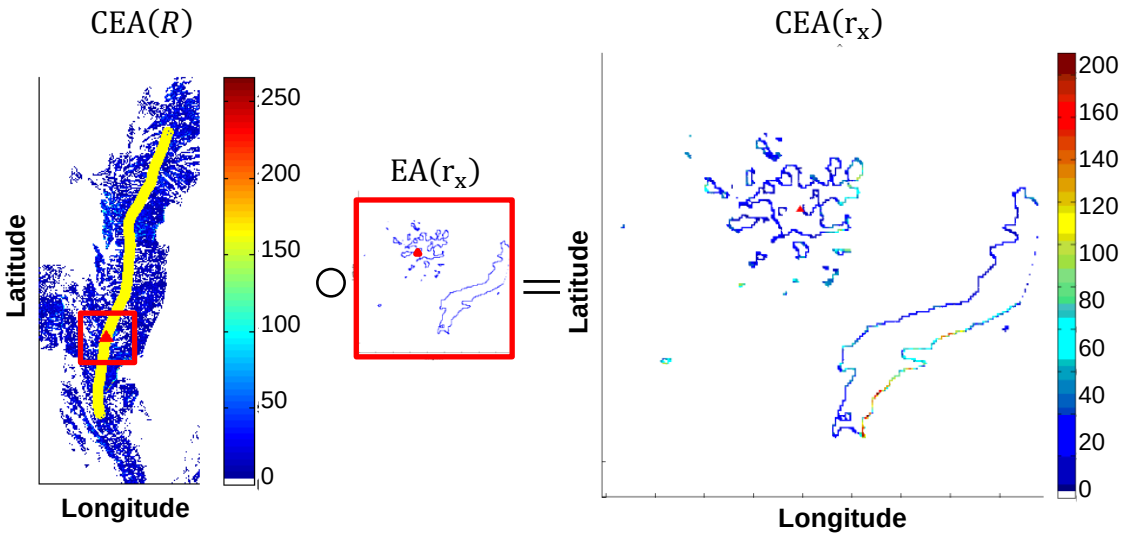



**Figure 10. a) *CEA($r_x$)* is the Hadamard product of *CEA(R)* and *EA($r_x$)*.**

Figure 9.a shows a cumulative escape adjacency map for an 800-point route. The callout, outlined in red in Figure 9.b, is the escape adjacency map for a single $r_x$.

Interpretation of $CEA(R)$ is straightforward. The value of each pixel in the $CEA(R)$ map is the total number of $r_x \in R$ for which that pixel is escape adjacent. In Figure 9.a, the maximum radial length used to calculate the viewshed was 2500 meters. So, a point (or pixel) $p$ with a value of 80 is escape adjacent for 80 points ($r_x \in R$) along the route ($R$), all of which are within 2500 meters of $p$.

Once $CEA(R)$ has been assembled, cumulative escape adjacency for individual points along the road, $CEA(r_x)$, can be determined. $CEA(r_x)$ is found by taking the Hadamard product (entrywise product) of $EA(r_x)$ and $CEA(R)$.

$$CEA(r_x) = CEA(R) \circ EA(r_x) \tag{15}$$

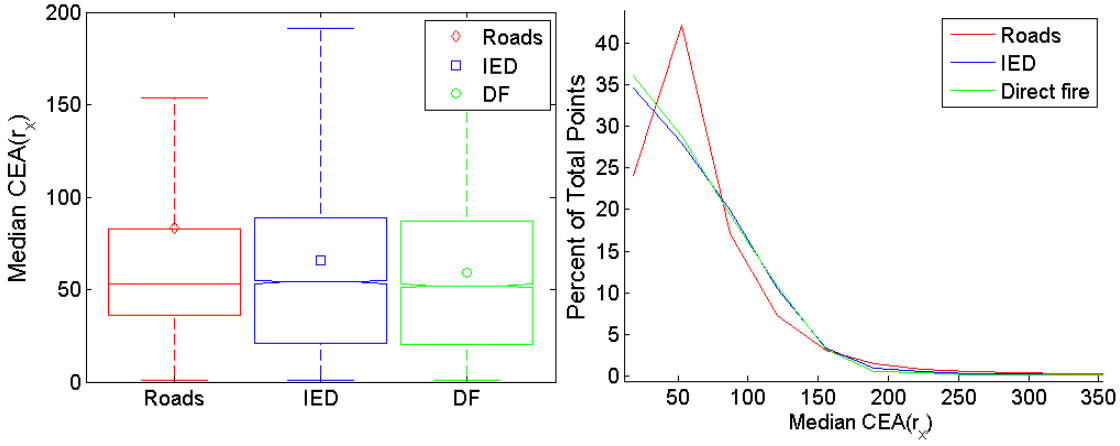



**Figure 11. Median *CEA($r_x$)*.**

Figure 10 shows how $CEA(R)$ and $EA(r_x)$ are multiplied to get $CEA(r_x)$. The result is a false-color map centered on $r_x$ showing a set of points that are escape adjacent with $r_x$ and with other points in $R$, where the color of a pixel depicts the number of $r_x \in R$ for which that pixel is escape adjacent.In the figure, the escape adjacent points colored red can see $r_x$ and approximately 200 other points along $R$.

*Viewshed Feature: Median Cumulative Escape Adjacency*

$CEA(r_x)$ provides a mechanism for describing the visibility and escape adjacency for a particular $r_x$. This allows an attacker or a target to determine the points that are likely to provide good support to a conflict event. Locations with visibility and a high CEA value are very useful for monitoring and overwatch. As previously noted, locations (pixels) with a value greater than one are escape adjacent both for $r_x$ as well as for other points. Median cumulative escape adjacency provides some indication of how visible $r_x$ and other points in $R$ are to surrounding escape adjacent locations and provides and ability to roughly assess the degree of conflict event support available to $r_x$.

$$\widetilde{CEA(r_x)} = median\big(CEA(r_x)\big) \quad (16)$$

Figure 11 shows the distribution of $\widetilde{CEA(r_x)}$. It appears that both types of events and roads are all drawn from similar distributions. Maximum and minimum $\widetilde{CEA(r_x)}$ can be found in Appendix B.3.

*Optimal Cumulative Escape Adjacency*

$CEA(r_x)$ provides the set of all points that are escape adjacent for some particular $r_x$ and may be escape adjacent for other $r_x \in R$. However, perusal of this exhaustive list of potential Monitor/Control locations is probably not common for risk averse actors and probably not representative of actual human behavior. Instead, a human seeking a good Monitor/Control location probably selects a general area with good potential sites and then selects the optimal location within or near that general area. This fits with our understanding of ”thin-slicing” as proposed by Gladwell [55] where decisions made are strongly influenced by intuition, instinct. This intuition or instinct is the outcome of an unconscious or subconscious integration of available facts and impressions.

Thus, in the search for a good location to support a conflict event, attackers may follow, consciously or unconsciously, a three-step process: (1) identify a general area that appears to meet Monitor/Control criteria; (2) move towards and around the selected general area; and (3) choose the locally optimal site at or near the selected general area for use as a Monitor/Control location. Optimal Cumulative Escape Adjacency attempts to capture the notion that humans are often willing to make some level of effort in order to improve their position or outcome. Assuming that a ‘better’ location has greater

cumulative escape adjacency, a reduced, more optimal set of $CEA(r_x)$ locations can be selected by discarding points that have nearby neighbors with greater escape adjacency.

Let *w* be the maximum distance that a human actor is willing to move in order to improve a position. Then,

$$CEA'(r_x)_{i,j} = \begin{cases} CEA(r_x)_{i,j} \mid CEA(r_x)_{i,j} \equiv max(mat\_w_{i,j}) \\ 0 \end{cases} \quad \forall\, i,j \in CEA(r_x) \quad (17)$$

, where $mat_w_{i,j}$is a circular window with radius *w* centered on $CEA(r_x)_{i,j}$. Each location in the resulting reduced set of locations can be considered locally optimal within a radius of *w* with respect to maximum cumulative escape adjacency.

*Viewshed Feature: Route Visibility*

For the tactics used in many conflict events, simple intervisibility with $r_x$ is not enough. Visibility of the approaches to $r_x$ is also important and the total extent of the visible area needed for a particular tactic varies with terrain and tactics. In the case where the target is mobile, an actor at a Control location may need sufficient visible extent to estimate vehicle speed accurately in order to trigger the IED or fire a weapon at a preselected site. Better roads and faster targets increase the total visibility required. Intervisibility over larger extents may also be required for the Control actor to ensure that the target is appropriate for the attack being planned. For example, an ambush using light shoulder-fired weapons should not engage a heavily armed patrol. In this situation, a Control actor might want to see all of the vehicles in a patrol before choosing to initiate the attack. Attack scale may also play an important role. A visible stretch of road may be required for a large-scale ambush. The Control actor is likely to want to have the entire

target patrol within the kill zone before initiation of an attack.

Let $w$ be the maximum extent of the approaches under consideration. Then $R'_w(r_x)$ is the subset of $R$ within distance $w$ of $r_x$.

$$R'_w(r_x) = R \mid distance(r_x, r_i) < w \qquad \forall \ \ i \ in \ R \tag{18}$$

Then the route visibility of the approaches to $r_x$ for $CEA'(r_x)_{i,j}$ is the fraction of total points in $R'_w(r_x)$ that have intervisibility with $CEA'(r_x)_{i,j}$.

$$Vis_w\big(CEA'(r_x)_{i,j}\big) = \frac{\sum_{i,j}\big(LOS\big(CEA'(r_x)_{i,j},\ R'_w(r_x)_k\big) \equiv 1\big)}{|R'_w(r_x)|} \qquad \forall\ k \ in \ R'_w(r_x) \tag{19}$$

, and the route visibility for all points in $CEA'(r_x)$ is the matrix

$$Vis_w\big(CEA'(r_x)\big) = \big[\, Vis_w\big(CEA'(r_x)_{i,j}\big)\big] \qquad i,j \ \in \ CEA'(r_x) \tag{20}$$

Once $Vis_w\big(CEA'(r_x)\big)$ has been calculated, then the median route visibility for the *CEA* points surrounding $r_x$ is the median of all values in $Vis_w\big(CEA'(r_x)\big)$

$$\widetilde{Vis_w\big(CEA'(r_x)\big)} = median\left(Vis_w\big(CEA'(r_x)\big)\right) \tag{21}$$

Figure 12 shows the importance of route visibility for approaches of 250 meters. In the figure, conflict events are significantly more likely to have better median route visibility than typical road points. In fact, the boxplot shows that a majority of IED and DF sites have approaches that are very exposed or visible (> 50%). In the histogram, note that approximately 30% of all conflict events have greater than 95% route visibility for all road points within 250 meters.

Appendix B.4 shows median route visibility at radii of 100, 500, and 1000 meters. Appendices B.5 and B.6 show maximum and minimum route visibility at radii of 100, 250, 500, 1000 meters.

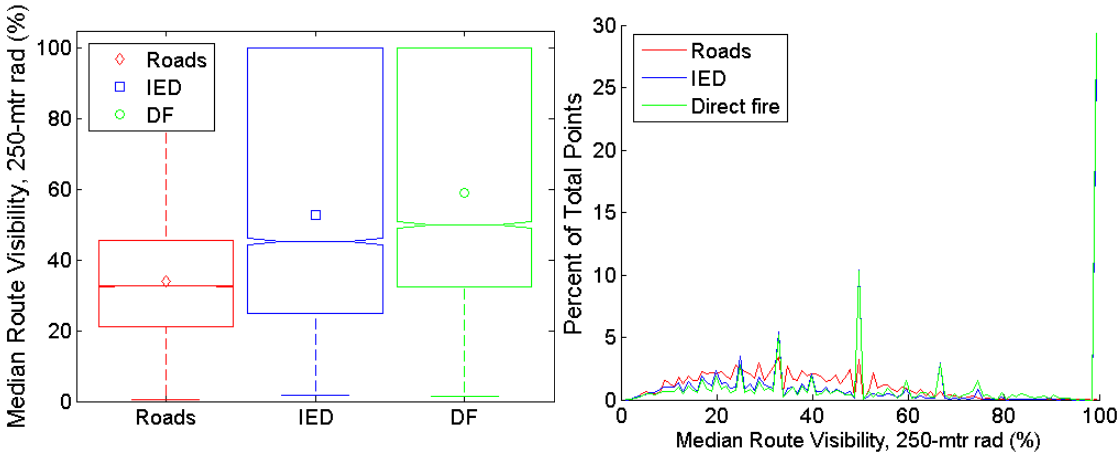



**Figure 12. Median route visibility at 250 meters.**

## Sparse Viewshed

In some cases, analysis of the exhaustive viewshed fails to capture salient features. The noisiness of discrete data may hide general trends over larger areas. Also, the use of symmetric windows centered on $r_x$ may conceal or wash out interesting asymmetric features. In these cases, sparse viewshed provides a mechanism for feature extraction that summarizes or constrains key viewshed features at scales appropriate to the feature being analyzed.

Sparse viewshed models terrain in a way that might be similar to the mental model constructed by a human assessing terrain. Humans tend to assess terrain by taking notice of major features, like hilltops, ridgelines, and running water. A mental model is constructed that locates these features in relation to each other. When a specific task needs to be accomplished, a human might also notice smaller or more specific features. For example, a hiker will notice the slope and roughness of possible routes. Sparse

viewshed provides a mechanism to simplistically model the limits of visibility. These limits can be used to estimate viewshed and to build viewshed-constrained versions of several common geomorphometric features.

To build a sparse viewshed, calculate viewshed as described in Equations 7-9. However, modify Equation 8 to include a stopping critiera, *tol,* and select an $N_s$ appropriate for the feature being collected. Define *tol* to be the maximum number of consecutive points along a radial that may have LOS=0. The end of the radial is set to be the last visible point before *tol* is exceeded. Once *tol* is exceeded, all more distant points are set to zero.

$$LOS(L_i) = \left[slope\left(l_{i,0}, l_{i,j}\right) \geq max\left(slope\left(l_{i,0}, l_{(i,1),\dots,(i,j-1)}\right)\right)\right], for\ j = 1, \dots, m \quad (22)$$

$$while\ \left(\sum_{k=j-tol}^{j} LOS\left(L_{i,k}\right)\right) > 0$$

The inclusion of the stopping criteria provides a mechanism to determine the length of a mostly uninterrupted sightline along a particular azimuth. In other words, all or almost all points between $r_x$ and the end of a radial are visible. The first significant gap in intervisibility occurs beyond the end of the radial. The *tol* variable is used to specify the width of a gap considered significant. Once the stopping criterion has been incorporated, sparse viewshed is calculated as described in Equation 9. The choice of $N_s$ is a tunable parameter that varies with the degree of summarization desired.

Figure 13 shows a sparse viewshed overlaying an exhaustive viewshed. The green pixels in the figure are an exhaustive viewshed calculated using a radial sweep algorithm with large $N_s$ as described in Equations 7-9. A sparse viewshed is formed

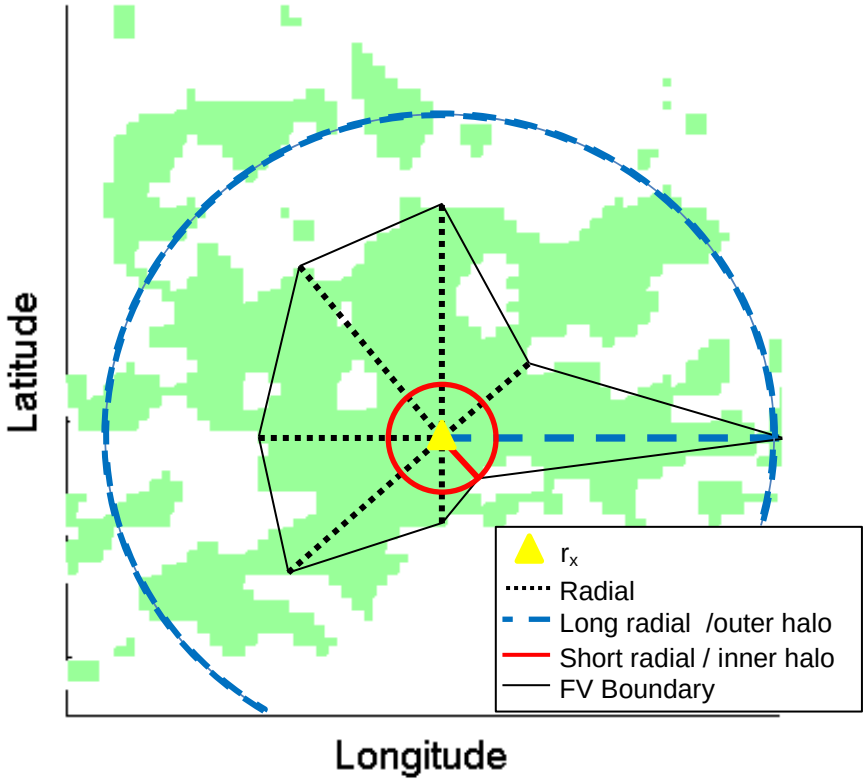



**Figure 13. Example of a sparse viewshed using eight radials.**

using $N_s = 8$ evenly spaced radials with the ends of the radials being determined as described in Equation 22. The *tol* criterion was set to a value of 2, so the radials stop when a gap of three or more pixels is encountered.

The thin black line joining the ends of the radials denotes the sparse viewshed boundary. In the figure, the longest radial is marked with a heavy-dash blue line. A heavy-dash line marks a circle whose radius is the longest radial. A solid red line marks a circle whose radius is the shortest radial. These two circles form the outer and inner boundary of a Halo or annulus. Key elements of conflict events planned by risk-averse attackers will occur within this Halo. Elements closer to $r_x$ than the red circle are likely

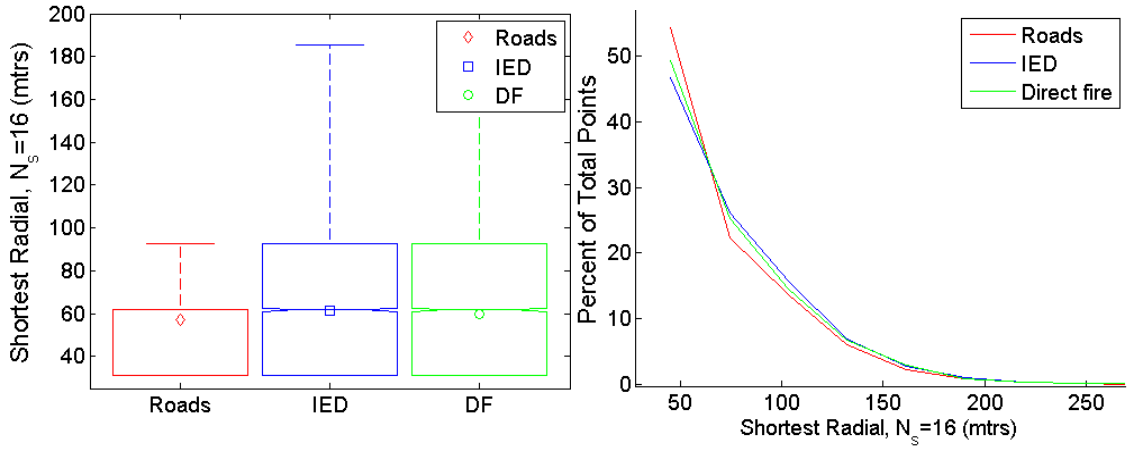



**Figure 14. Distribution of shortest radial, for sparse viewshed with $N_s$ = 16.**

to be exposed to view by the target. Elements further away from $r_x$ than the blue circle are likely to have poor or no visibility of $r_x$. Thus, analysis of terrain within this Halo may provide key insights in attacker tactics and potential use of terrain.

Several features can be calculated from sparse viewshed.

*Viewshed Feature: Shortest Radial*

The length of the shortest radial denotes the nearest location along a selected radial where there is a significant gap in intervisibility. It is also an estimate of the length of the shortest sightline. For some types of events, the shortest radial may describe the closest place to $r_x$ where attackers can conceal themselves. An interruption in intervisibility captured by the shortest radial may also be an indication that nearby terrain changes abruptly.

Figure 14 shows the distribution of the length of the shortest radial when $N_s$ = 16. Although all conflict events seem to share a common distribution, they are clearly

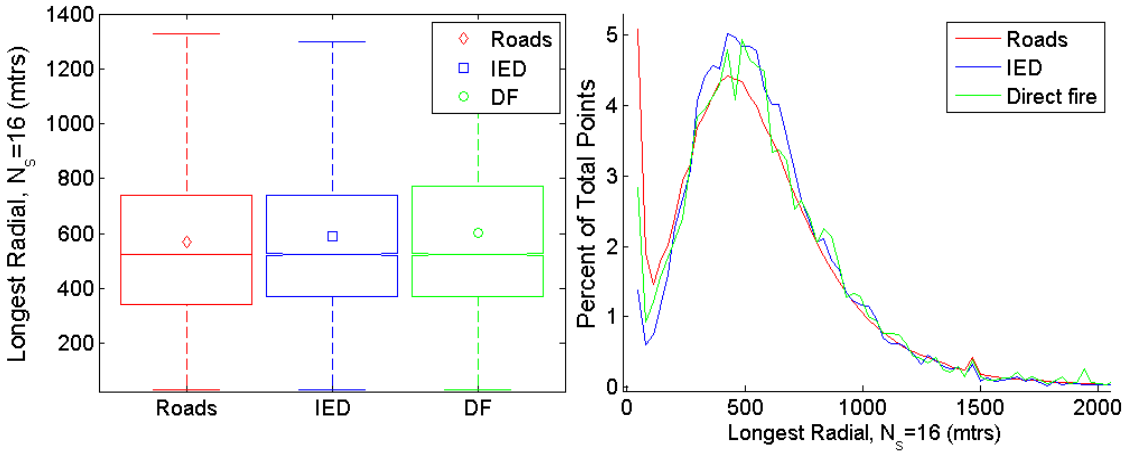



**Figure 15. Distribution of longest radial, for sparse viewshed with $N_s$ = 16.**

distinct from a majority of road points. The shortest radial for various sizes of $N_s$ (4, 8, 32, and 64 radials) can be found in Appendices B.7.

*Viewshed Feature: Longest Radial*

The length of the longest radial is the length of the longest uninterrupted or mostly uninterrupted sightline. For some types of conflict events, the longest radial may describe the direction in which Monitor actors may possibly be found. A long radial may often highlight linear features that lie along a radial, like a river valley or ridge, or long gentle slopes, where intervisibility is not interrupted. Also, long radials tend to indicate that the terrain along that radial tends to be relatively smooth.

Figure 15 shows the distribution of the longest radial. It appears that conflict events and roads share a common distribution. The longest radial for various sizes of $N_s$ (4, 8, 32, and 64 radials) can be found in Appendices B.8.

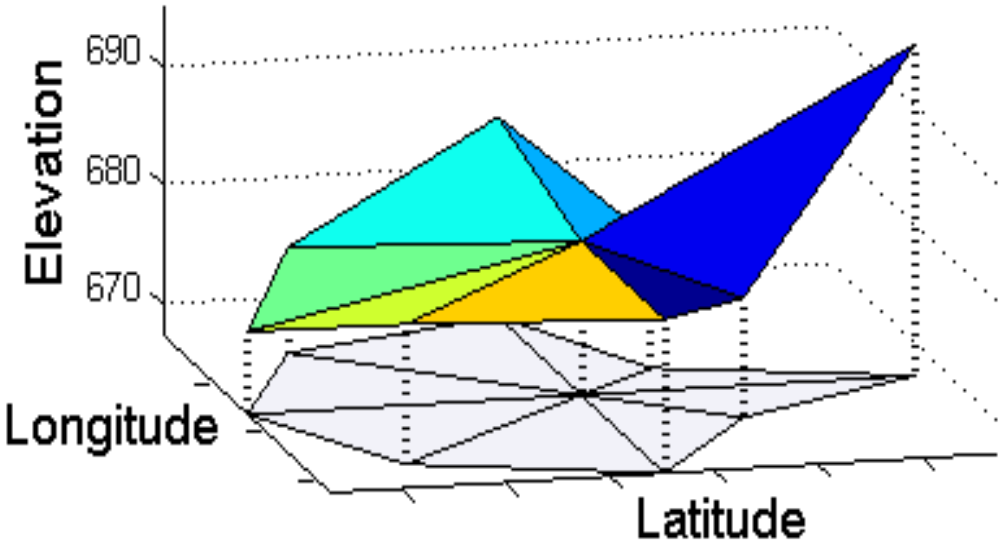



**Figure 16. Sparse viewshed portrayed in three dimensions.**

*Viewshed Feature: Local Openness*

Local openness quantifies the general lay of the land as described by the radials of a sparse viewshed. Defined as the mean of the slopes of the radials, upward openness provides some insight into the general shape of the terrain. Smaller values are found on flatter terrain, which tends to be more open, while larger values are found in more rugged terrain.

Let $abs\big(slope_i(r_x)\big)$ be the absolute value of the slope between $r_x$ and the end of radial $i$. Then local openness for a sparse viewshed is calculated as

$$\mu_{slope}(r_x) = \frac{1}{N_s}\sum abs\big(slope_i(r_x)\big) \qquad for\ i = 1, \dots, N_s \qquad (23)$$

Figure 16 uses eight radials to depict a sparse viewshed in three dimensions. The colored triangles estimate the terrain's surface between adjacent radials. The gray area

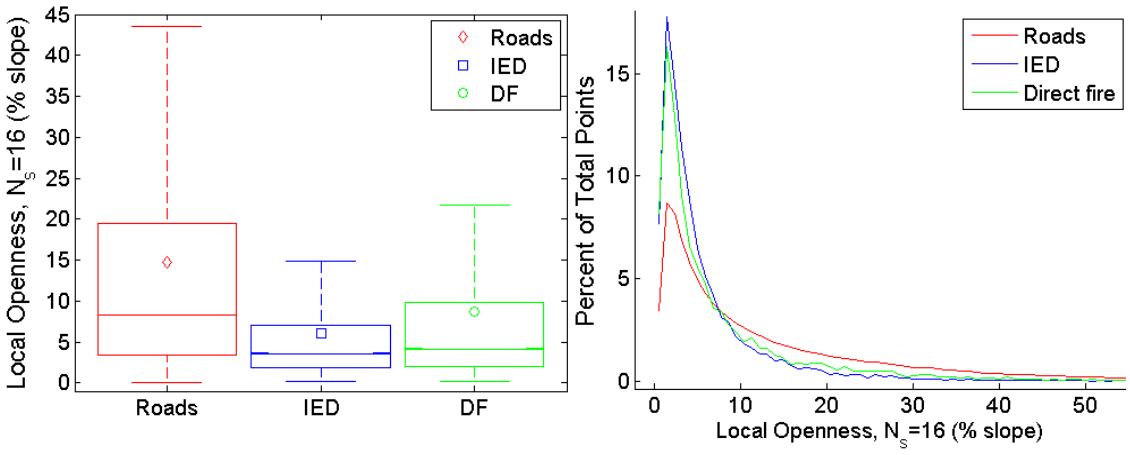



**Figure 17. Distribution of local openness, for sparse viewshed with $N_s$ = 16.**

underneath represents the planimetric area described by the radials. Local openness for this eight-radial sparse viewshed is the average of the absolute values of the slopes of the radials.

Figure 17 depicts local openness calculated using a sparse viewshed with 16 radials. Interestingly, conflict event sites tend to be more open than most road sites with IED sites distributed across the smallest range of openness. Local openness for various sizes of $N_s$ (4, 8, 32, and 64 radials) can be found in Appendices B.9.

### Geomorphometric Features

There is a wide variety of geomorphometric parameters that describe the underlying morphographic structure of terrain. For a baseline, we use the three part geometric pattern proposed by Iwahashi and Pike in [22] based in part on the work of Horn [133] . The pattern, consisting of slope gradient, texture and local convexity, is designed to capture key differentiating features of different landscapes. Other geomorphometric features are also collected.

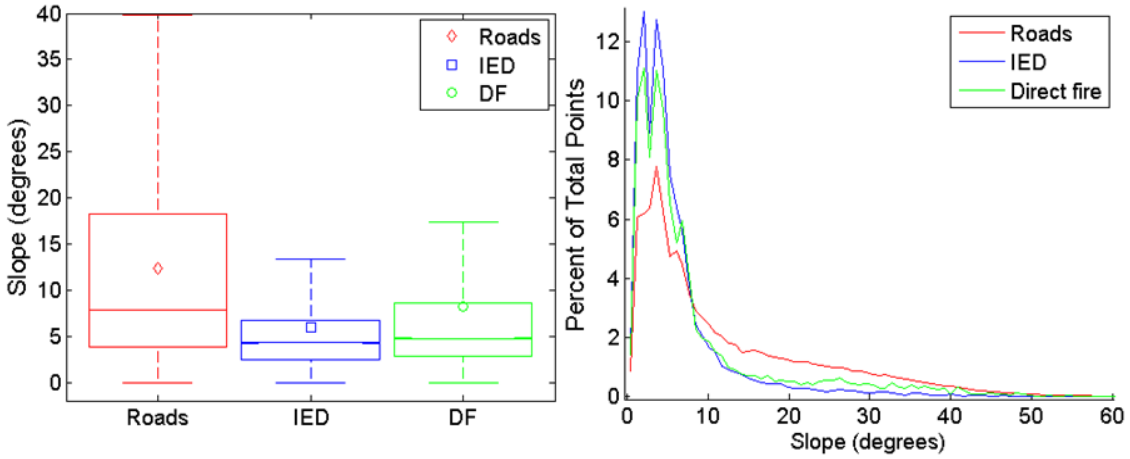



**Figure 18. Distribution of slope.**

Three important facts should be noted regarding the collection of geomorphometric features in this research. First, the resolution of the elevation data is fixed and limited to approximately 30 meters. Since terrain features are strongly scale dependent, this resolution may be insufficient to capture larger features and vice versa. Second, the geomorphometric features selected are representative and commonly found in the literature. However, they may not be optimal for this resolution or terrain type. Finally, some parameterization, especially for window sizes, is strongly based on anecdotal estimation of the distances required for certain asymmetric warfare activities. These estimates are likely to change if field-based analysis becomes available.

*Feature: Slope*

Slope is defined as the change in elevation per meter of distance along the path of steepest ascent or descent. It is calculated using a 3-pixel x 3-pixel (3x3) analysis window centered on the elevation map pixel containing the location of interest, $r_x$. Matlab provides slope as an output of the *gradientm* function and calculates it as:

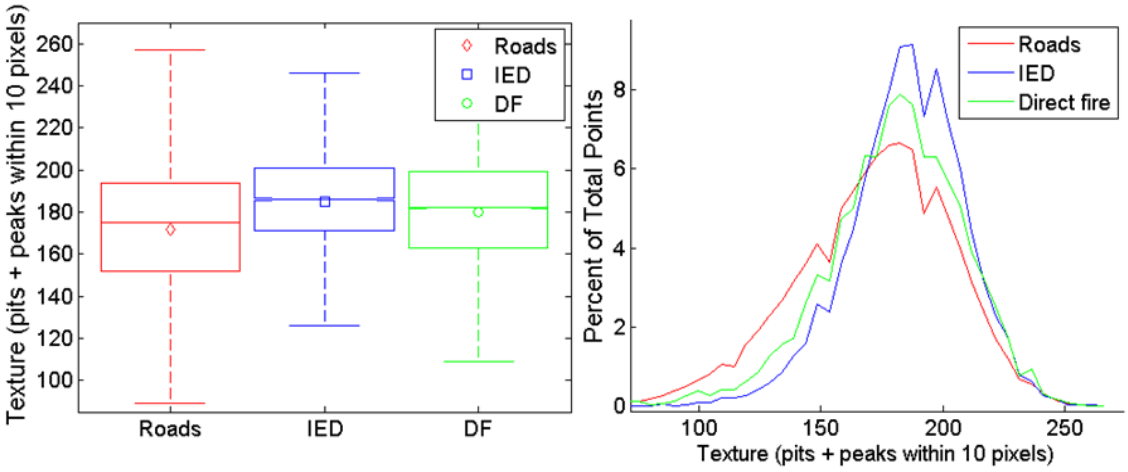



**Figure 19. Distribution of texture, as defined by Iwahashi and Pike.**

$$\nabla\bar{Z} = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right), \qquad SLOPE = arctan(|\nabla\bar{Z}|) \tag{24}$$

Figure 18 compares the distributions of slope for roads, IED, and DF. Notably, events tend to be on flatter sites than most roads. This is similar to the observations from the local openness feature. Additionally, conflict events and roads appear to be from different populations, based on the width and positioning of the boxplot notches.

*Feature: Texture*

Texture is defined by Iwahashi and Pike as the total number of pits and peaks within a ten pixel radius of a point [22]. A 3-pixel x 3-pixel (3x3) median filter is used to smooth the original digital elevation model (DEM). The output of the filter is a smoothed DEM that is subtracted from the original DEM and examined for magnitude. Magnitudes greater than zero indicate peaks and magnitudes less than zero indicate pits.

$$T = |DEM - f(DEM)| > 0, \quad f\ is\ a\ median\ filter \tag{25}$$

$$texture(r_x) = \sum_{m,n} T_{m,n} \mid distance(r_x, T_{m,n}) \leq 10\ pixels \tag{26}$$

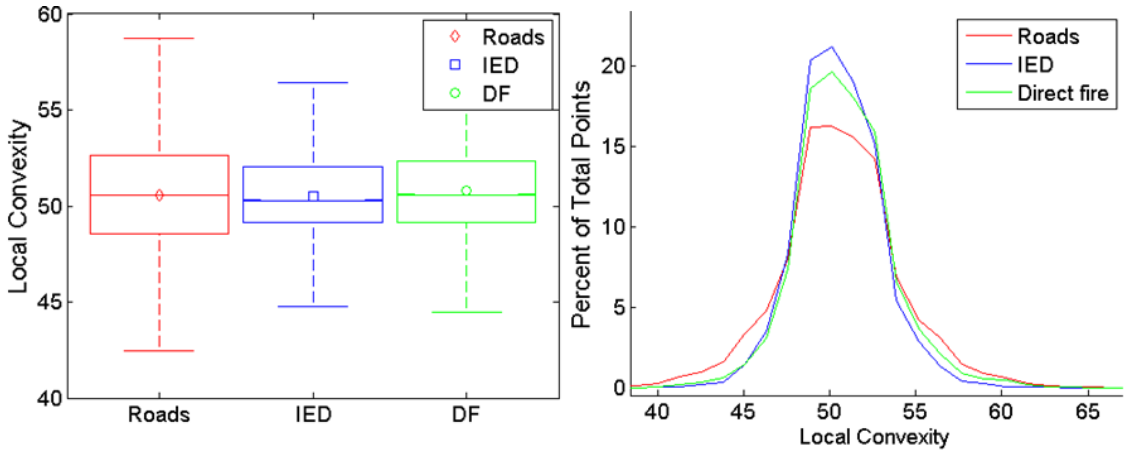



**Figure 20. Distribution of local convexity, as defined by Iwahashi and Pike.**

For context, in the case of Afghanistan, where the publically available DEMs have a resolution of approximately 30 meters, individual textures are calculated over a circular area encompassing approximately 280,000 meters$^2$. Figure 19 compares the distributions of texture for roads, IED, and DF events. While all three classes follow a similar distribution, conflict events tend to have greater texture than roads. The boxplots indicate that all three classes are probably drawn from different populations.

*Feature: Local Convexity*

The convexity of individual map pixels is found by calculating the surface curvature of a 3x3 DEM subgrid using a Laplacian filter. Matlab calculates convexity, *Conv*, using a convolution kernel *K*,

$$K = \begin{bmatrix} 0.1667 & 0.6667 & 0.1667 \\ 0.6667 & -3.3333 & 0.6667 \\ 0.1667 & 0.6667 & 0.1667 \end{bmatrix} \tag{27}$$

$$Conv_{i,j} = \sum_{k=1}^{3}\sum_{l=1}^{3} DEM\,(i+2-k, j+2-l)K(k,l) \qquad \forall\, i,j \in DEM \tag{28}$$

Since local convexity, as defined by Iwahashi and Pike, only counts pixels with positive values, all values less than or equal to zero can be set to zero.

$$C = Conv > 0 \tag{29}$$

Then, local convexity is defined as the percentage of convex upward (positive) pixels within a ten pixel radius of a point [22].

$$LocalConvexity(r_x) = \frac{100}{N_p}\sum_{m,n} C_{m,n} \mid distance(r_x, C_{m,n}) \leq 10\ pixels \tag{30}$$

, where $N_p$ is the number of pixels within a ten pixel radius of $(i,j)$.

As shown in Figure 20, for all types of conflict events, the distribution of local convexity varies little and the range of convexity values is relatively narrow.

*Feature: Elevation Range*

Elevation range is the difference between the highest and lowest elevation in a window [43]. For this analysis, we examine the difference at a radius of 350 meters. Note that in Equations 31 and 22, $G(r_x)_{rad}$ is a matrix of elevations within $rad$ meters of $r_x$.

$$G(r_x)_{rad} = DEM \mid distance(r_x, DEM_{m,n}) < rad \quad \forall\, m,n \in DEM \tag{31}$$

$$range(r_x)_{rad} = max(G(r_x)_{rad})) - min(G(r_x)_{rad}) \tag{32}$$

Figure 21 shows that the elevation range is markedly different between roads and conflict events using a window size of 350 meters. At this window size, roads have a larger range of values while conflict events are seen on terrain with smaller ranges. This small range of values may indicate that attackers prefer flatter or smoother ground in the vicinity of an attack site.

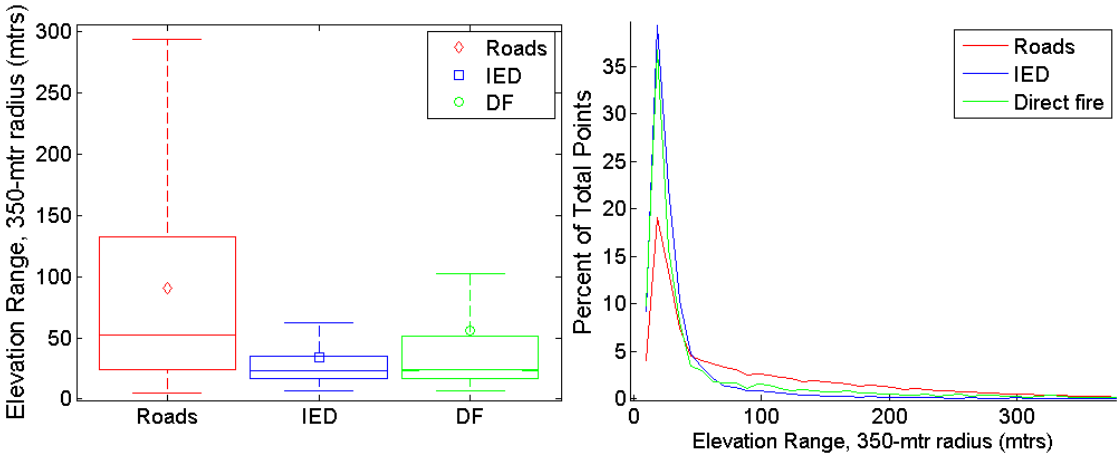



**Figure 21. Distribution of elevation range at a radius of 350 meters.**

Graphs showing elevation range at window radii of 50, 100, 500 and 1000 meters can be found in Appendix B.10.

*Feature: Roughness*

Roughness uses the standard deviation of elevation across a window to estimate the texture of a surface. Large standard deviations are an indication of a more undulating or rougher surface. Determine $G(r_x)_{rad}$ as in Equation 31. Then the standard deviation of elevation across $G(r_x)_{rad}$ is

$$\sigma(r_x)_{rad} = \sigma(G(r_x)_{rad}) \tag{33}$$

, where σ is the standard deviation function.

Figure 21 compares the distribution of roughness across the three classes for a window size of 350 meters. The classes of events tend to be on smoother ground than roads. Notably, the distribution of roughness closely resembles the distribution of elevation range. Appendix B.11 contains figures for roughness across other window sizes.

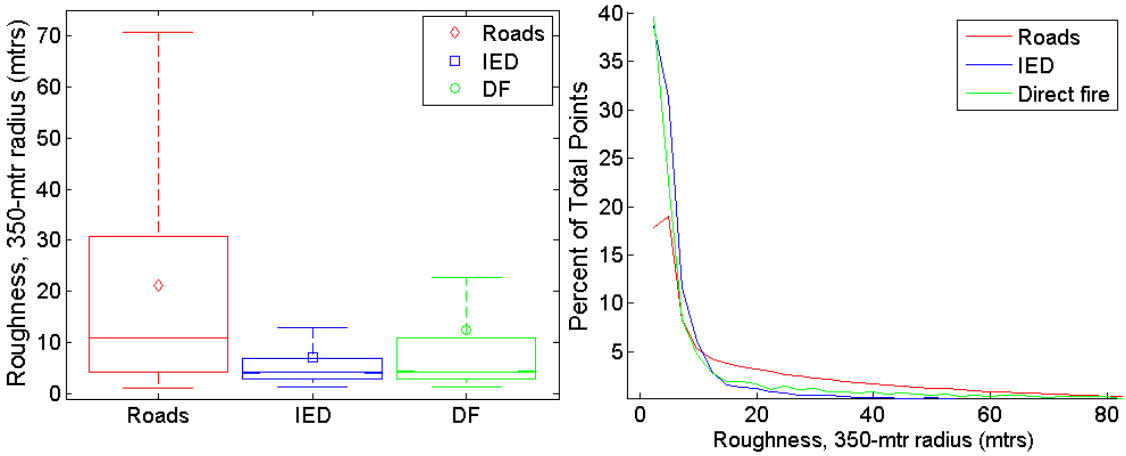



**Figure 22. Distribution of roughness over a radius of 350 meters.**

## Social/Cultural Features

Some features are not directly linked to geomorphometry or viewshed. These features, often related to social or cultural factors, capture aspects of site selection not related to the land itself. Proximity to populated areas is explored below.

### *Social/Cultural Feature: Proximity to Populated Areas*

In some cases, attackers may require access to populated areas. Access may be for logistical reasons, e.g. attackers need access to communications, lodging, etc., or for cover and concealment, e.g. attackers may be able to hide within the local populace.

Figure 23 shows the distribution of the distance from conflict event sites to the nearest population center with more than 1000 inhabitants. Notably, conflict events tend to be much closer to inhabited areas than points along roads with a median value of approximately 1 km. Appendix B.13 shows the distributions for distances to populated areas ranging in size from 1 to 1 million inhabitants.

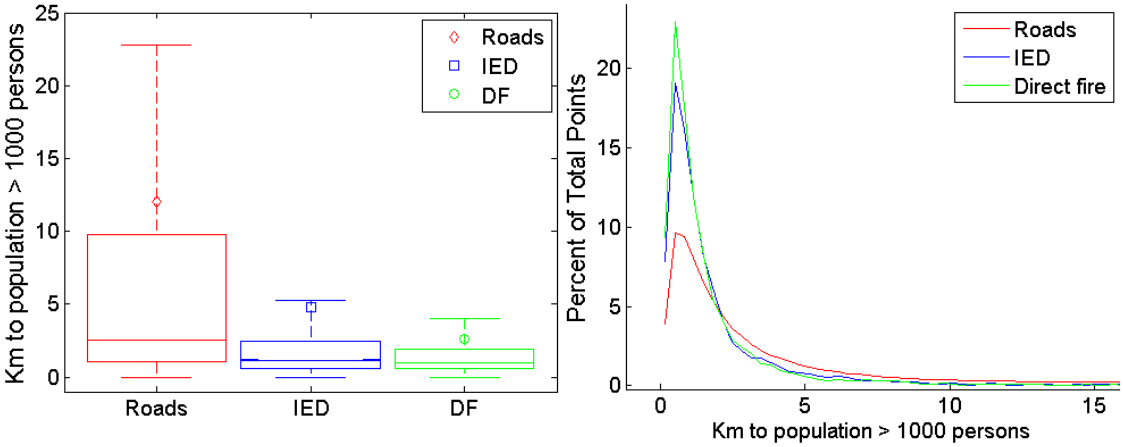



**Figure 23. Distribution of distance to nearest populated area with greater than 1000 inhabitants.**

## Other Features

Several other features are presented in Appendix B. These features are similar to those already presented here and are included in the appendix for completeness:

- Sparse Viewshed Mean Radial Length is presented in Appendix B.12.
- Sparse Viewshed Planimetric Area is presented in Appendix B.14.
- Sparse Viewshed Rugosity is presented in Appendix B.15.
- Sparse Viewshed Shape Complexity is presented in Appendix B.16.

## Comparison of Features across Classes

In the previous sections, a total of 20 different features were collected. Some of the features were collected at various geographic window sizes or at various resolutions. The resulting feature set consists of 77 measurements. Table 2 summarizes the results of multiple Kruskal-Wallis tests to determine if the observations for distinct classes (roads, IED and DF) come from different populations.

The Kruskal-Wallis (KW) test is a non-parametric statistical test designed to assess if the measurements for two or more classes come from the same population. It tests the null hypothesis —that all measurements are drawn from the same population— by comparing the medians for each class. As a non-parametric test, KW makes no assumptions about distributions of the measurements or residuals. The test only assumes that all measurements are independent and that they are all drawn from the same continuous distribution. The output of the KW test is the *p*-value for the null hypothesis. Small *p*-values call into question the null hypothesis and indicate that, in the measurements used, at least one class median is significantly different from the others.

Table 2 provides the output of multiple KW tests. For each of the 77 measurements in the two-column table, three different KW tests were run comparing (1) roads with IED events, (2) roads with DF events, and (3) IED events with DF events. The *p*-values are captured in the table and significant values ($p <= 0.05$) are highlighted in yellow. As shown, roads and events (of any type) are very likely to have been drawn from different populations. These results hold for IED versus roads in 73 of 77 features. DF versus roads shows similar results. Interestingly, IED and DF events appear to come from the same population for 55 of 77 measures. For IED and DF events, the 22 measures assessed by the KW test to come from different classes actually include 11 measures of route visibility and five measures of shape complexity. In other words, IED and DF events are assessed to be from different populations in only eight of 20 features.

Appendix C contains tables summarizing key statistics for datasets used in this research (mean, standard deviation, skewness, kurtosis, maximum and minimum).

**Table 2. Kruskal-Wallis test results**

| Feature | Rds : IED | Rds: DF | IED : DF |
|---|---|---|---|
| Elevation | 1.11E-13 | 1.32E-12 | 0.003928 |
| Slope | 5.43E-255 | 3.05E-122 | 1.08E-06 |
| Convexity (IW) | 0.5576 | 0.06056 | 0.04094 |
| Texture (IW) | 2.96E-108 | 4.44E-56 | 0.005387 |
| Elev. Rng (50) | 1.13E-201 | 2.94E-146 | 0.7953 |
| Elev. Rng (100) | 8.10E-239 | 8.89E-150 | 0.0926 |
| Elev. Rng (350) | 5.935e-315 | 4.60E-188 | 0.1276 |
| Elev. Rng (500) | 4.397e-322 | 1.88E-198 | 0.4111 |
| Elev. Rng (1000) | 1.808e-316 | 5.96E-192 | 0.5589 |
| Roughness (50) | 9.85E-196 | 1.53E-143 | 0.6819 |
| Roughness (100) | 1.07E-223 | 5.28E-144 | 0.2267 |
| Roughness (350) | 8.66E-300 | 1.43E-179 | 0.1431 |
| Roughness (500) | 0.00E+00 | 1.14E-191 | 0.5112 |
| Roughness (1000) | 0.00E+00 | 2.12E-196 | 0.7572 |
| Vis Idx(100_350) | 1.54E-05 | 4.80E-08 | 0.1539 |
| Vis Index (350) | 9.44E-06 | 3.38E-08 | 0.1591 |
| Vis Index (500) | 0.0001491 | 1.93E-08 | 0.07718 |
| Vis Index (1000) | 0.1004 | 2.67E-06 | 0.02136 |
| SCID (100_350) | 0.01643 | 1.11E-05 | 0.06298 |
| SCI D (350) | 9.44E-06 | 3.38E-08 | 0.1591 |
| SCI D (500) | 0.0001491 | 1.93E-08 | 0.07718 |
| SCID (1000) | 0.1004 | 2.67E-06 | 0.02136 |
| Short radial (4) | 1.47E-05 | 1.78E-05 | 0.6582 |
| Long radial (4) | 1.18E-07 | 7.73E-07 | 0.7779 |
| Mean radial (4) | 1.07E-09 | 2.67E-10 | 0.3949 |
| Local open (4) | 2.03E-237 | 1.02E-147 | 0.1425 |
| Planimet area (4) | 1.65E-09 | 5.09E-10 | 0.4079 |
| Rugosity (4) | 4.60E-07 | 0.0004824 | 0.4863 |
| Shape cmplx (4) | 7.81E-25 | 1.86E-10 | 0.03742 |
| Short radial (8) | 7.21E-07 | 2.55E-07 | 0.5359 |
| Long radial (8) | 8.77E-13 | 1.46E-09 | 0.8917 |
| Mean radial (8) | 2.04E-15 | 2.87E-12 | 0.8934 |
| Local open (8) | 2.55E-241 | 9.74E-149 | 0.1198 |
| Planimet. area (8) | 4.49E-14 | 1.09E-11 | 0.8191 |
| Rugosity (8) | 2.86E-10 | 3.06E-09 | 0.7138 |
| Shape cmplx (8) | 2.02E-20 | 1.08E-09 | 0.1107 |
| Short radial (16) | 1.28E-07 | 0.000156 | 0.5149 |
| Long radial (16) | 3.78E-19 | 6.75E-13 | 0.6864 |
| Mean radial (16) | 1.22E-18 | 1.20E-14 | 0.8745 |

| Feature | Rds : IED | Rds: DF | IED : DF |
|---|---|---|---|
| Local open (16) | 9.07E-239 | 3.29E-147 | 0.1321 |
| Planimet area (16) | 6.31E-18 | 1.55E-14 | 0.8096 |
| Rugosity (16) | 5.23E-18 | 6.57E-11 | 0.5024 |
| Shape cmplx (16) | 2.45E-20 | 1.63E-06 | 0.009604 |
| Short radial (32) | 1.12E-08 | 1.52E-07 | 0.8531 |
| Long radial (32) | 1.65E-20 | 1.02E-15 | 0.9818 |
| Mean radial (32) | 2.86E-19 | 1.04E-16 | 0.6169 |
| Local open (32) | 2.49E-240 | 2.44E-147 | 0.1145 |
| Planimet area (32) | 1.38E-18 | 1.64E-16 | 0.5914 |
| Rugosity (32) | 1.23E-21 | 6.36E-11 | 0.1907 |
| Shape cmplx (32) | 3.12E-14 | 0.001562 | 0.00719 |
| Short radial (64) | 7.34E-10 | 7.76E-09 | 0.7676 |
| Long radial (64) | 2.74E-21 | 5.90E-18 | 0.7203 |
| Mean radial (64) | 1.40E-19 | 9.99E-17 | 0.6435 |
| Local open (64) | 8.24E-241 | 3.98E-148 | 0.1272 |
| Planimet area (64) | 6.62E-19 | 8.63E-17 | 0.5935 |
| Rugosity (64) | 5.19E-18 | 2.69E-14 | 0.9133 |
| Shape cmplx (64) | 3.88E-11 | 0.00224 | 0.03802 |
| Dist. to ppl (1) | 0 | 5.17E-246 | 0.6226 |
| Dist. to ppl (1k) | 0 | 1.61E-293 | 0.4416 |
| Dist. to ppl (10k) | 3.62E-182 | 6.51E-163 | 0.5654 |
| Dist. to ppl (50k) | 1.39E-246 | 1.16E-179 | 0.05796 |
| Dist. to ppl (100k) | 1.61E-23 | 1.58E-39 | 2.73E-08 |
| CEA min | 2.23E-79 | 1.17E-74 | 0.2858 |
| CEA max | 3.25E-168 | 2.18E-116 | 0.07424 |
| CEA med | 1.92E-248 | 4.56E-189 | 0.945 |
| Rte Vis. (min 1k) | 3.26E-127 | 4.17E-06 | 5.95E-26 |
| Rte Vis. (max 1k) | 0.0009066 | 0.7642 | 0.03552 |
| Rte Vis. (med 1k) | 0.3002 | 9.45E-09 | 0.000248 |
| Rte Vis. (min 500) | 1.04E-192 | 3.68E-17 | 3.47E-27 |
| Rte Vis. (max 500) | 5.61E-25 | 4.75E-09 | 0.01279 |
| Rte Vis. (med 500) | 7.91E-11 | 0.98 | 2.87E-05 |
| Rte Vis. (min 250) | 1.16E-190 | 1.54E-23 | 5.01E-20 |
| Rte Vis. (max 250) | 6.38E-54 | 3.41E-22 | 0.00094 |
| Rte Vis. (med 250) | 3.61E-41 | 1.21E-08 | 2.81E-05 |
| Rte Vis. (min 100) | 0.004383 | 1.49E-30 | 1.56E-11 |
| Rte Vis. (max 100) | 1.61E-15 | 5.41E-06 | 0.06075 |
| Rte Vis. (med 100) | 8.42E-40 | 0.05053 | 1.30E-10 |

## Summary

A total of 20 features were collected at various resolutions or using various geographic windows. These 20 features produced a total of 77 distinct measures. The following measures can be used to describe the Emplacement site:

- Elevation, slope, convexity, texture, elevation range at 50 meters, roughness at 50 meters, local openness (all resolutions), distance to populated areas, and route visibility at 100 meters.

Monitor/control sites can be described using:

- Elevation range and roughness at radii greater than 50 meters, visibility index, discrete shape complexity, long radial, short radial, mean radial, planimetric area, rugosity, sparse viewshed shape complexity, cumulative escape adjacency and route visibility at radii greater than 100 meters.

As seen by visual inspection of the boxplots, several of the measures are clearly different for different classes. Table 2 offers one way to quantify this difference and support the intuition gained from the visual inspection. Given this difference, it is probably possible to use these features for predictive analysis.

# CHAPTER V

# PREDICTIVE ANALYSIS OF ASYMMETRIC CONFLICT EVENTS

Predictive analysis uses a variety of techniques to analyze historical data for the purpose of making predictions about the future or about unvisited locations. In this chapter we propose algorithms to populate and use the MECH model. The purpose of this study is to design an accurate and robust classification algorithm that learns from available data under realistic constraints.

In the development of the algorithm, subset selection and principal component analysis are compared as mechanism to reduce dimensionality. Classification is performed on the resulting reduced model using supervised parametric and non-parametric techniques including support vector machines, discriminant analysis and nearest neighbor classifiers. Performance of these classifiers is compared in various situations including:

- under geographic constraints,
- under temporal constraints, and
- under combined geographic and temporal constraints

Constraint performance is assessed using classification error and examined as a function of sample size and sample density. Using the best-performing constraints as determined in the analysis described above, learners are proposed and performance is analyzed.

## The MECH Classification Algorithm

Conflict events are rare. When overlaid on a tokenized road map of Afghanistan, conflict events only occupy 0.8% of the total number of points comprising the roads. Conflict events are notably different than average road points. Route visibility (Figure 12), shortest radial (Figure 14), local openness (Figure 17), elevation range (Figure 21), and distance to populated areas (Figure 23) are all features where the distribution of conflict events and road points are clearly different. Knowing that this difference exists, we are able to design an effective classification algorithm that adaptively learns from the subset of MECH features most relevant for the area and time under consideration.

The MECH classification algorithm adapts to available data in several ways. First, normalization factors are dynamically determined from local data. This ensures that scaling is determined from the data set and appropriate for the terrain and tactics described by the data. Next, relevant features are determined from the same local data. Thus, the set of features identified as relevant varies with tactics, terrain, and time. Finally, parameterization of the model is derived from local data. Thus, each time the model is used for predictive analysis, it is uniquely tuned to historical events, terrain, and the tactics in use in the local area.

As previously noted in Chapter III, the MECH model is comprised of two tactical patterns: one composed of Emplacement features, $\tau_E(r_x)$, and one composed of Monitor/Control features, $\tau_H(r_x)$. Together, these features form the tactical pattern of the conflict event $T(r_x) = [\tau_E(r_x) \;\; \tau_H(r_x)]$.

This pattern forms the core of the MECH classification algorithm. A constrained set of points are identified that include both conflict event and non-conflict event sites. Features are collected from these sites and normalized. These normalized features are examined for relevance. The resulting reduced set is used to develop classification rules. Table 3 summarizes the MECH Classification Algorithm and provides an outline for the sections that follow.

**Table 3. The MECH Classification Algorithm**

| |
|---|
| **Objective**<br>Classify Points Along Roads |
| **Algorithm**<br>Step 1: Select a dataset.<br>Given a set of locations collected from conflict events and road points, select the conflict events and road points to include in the training set<br>(i) Apply geographic constraints<br>(ii) Apply temporal constraints<br>If there are enough events in the resulting constrained sample<br>(iii) Divide the data into training and test sets<br>(iv) Collect Emplacement and Monitor/Control (E and M/C) features<br><br>Step 2: Prepare the data<br>(v) Normalize the training set<br>(vi) Normalize the test set using scaling factors from the training set<br><br>Step 3: Train and assess the model<br>(vii) Select relevant features<br>(viii) Determine model parameters and/or hyperparameters<br>(ix) Learn the classification rules for the training set<br>(x) Estimate classification accuracy<br><br>Step 4: Classify<br>(xi) Apply the classification rules |

*Assessment and Experiment Construction*

In the following analyses, classification is performed using support vector machine (SVM), discriminant analysis (DA), and k-nearest neighbor (KNN) algorithms. Various filters or constraints are applied and the resulting datasets are classified using five-way cross-validation. Three classes of data are used.

Two historical event classes are formed by IED events within 100 meters of a known road ($IED_{100}$) and direct fire events within 100 meters of a known road ($DF_{100}$). Collectively, these two sets comprise $EVS_{100}$, the set of all events within 100 meters of a road. Roads are tokenized into discrete points at an interval of 30 meters, which coincides approximately with the elevation map resolution. Points along the road that are at least 250 meters from any conflict event ($RD_{250}$) are used as the non-event class. Many of the figures displayed are based on *ST*, a dataset consisting of 250 events drawn randomly from $IED_{100}$ and 250 events drawn randomly from $DF_{100}$. The dates and locations of the *ST* events provide a consistent set of geographic and temporal coordinates that are used to test and compare algorithms. Figure 24 shows the temporal and geographic distribution of *ST*.

The $RD_{250}$ dataset was chosen to minimize suspected problems with the conflict event data. First, it is not clear that the location of the conflict event is selected using consistent criteria. For example, the actual location of an exploded or discovered IED is very clear from physical observation. Blast marks, craters or the actual device can be seen and the location measured accurately. However, the reported location is not always so accurate. It may be the actual site of the explosion, an estimate of the location made

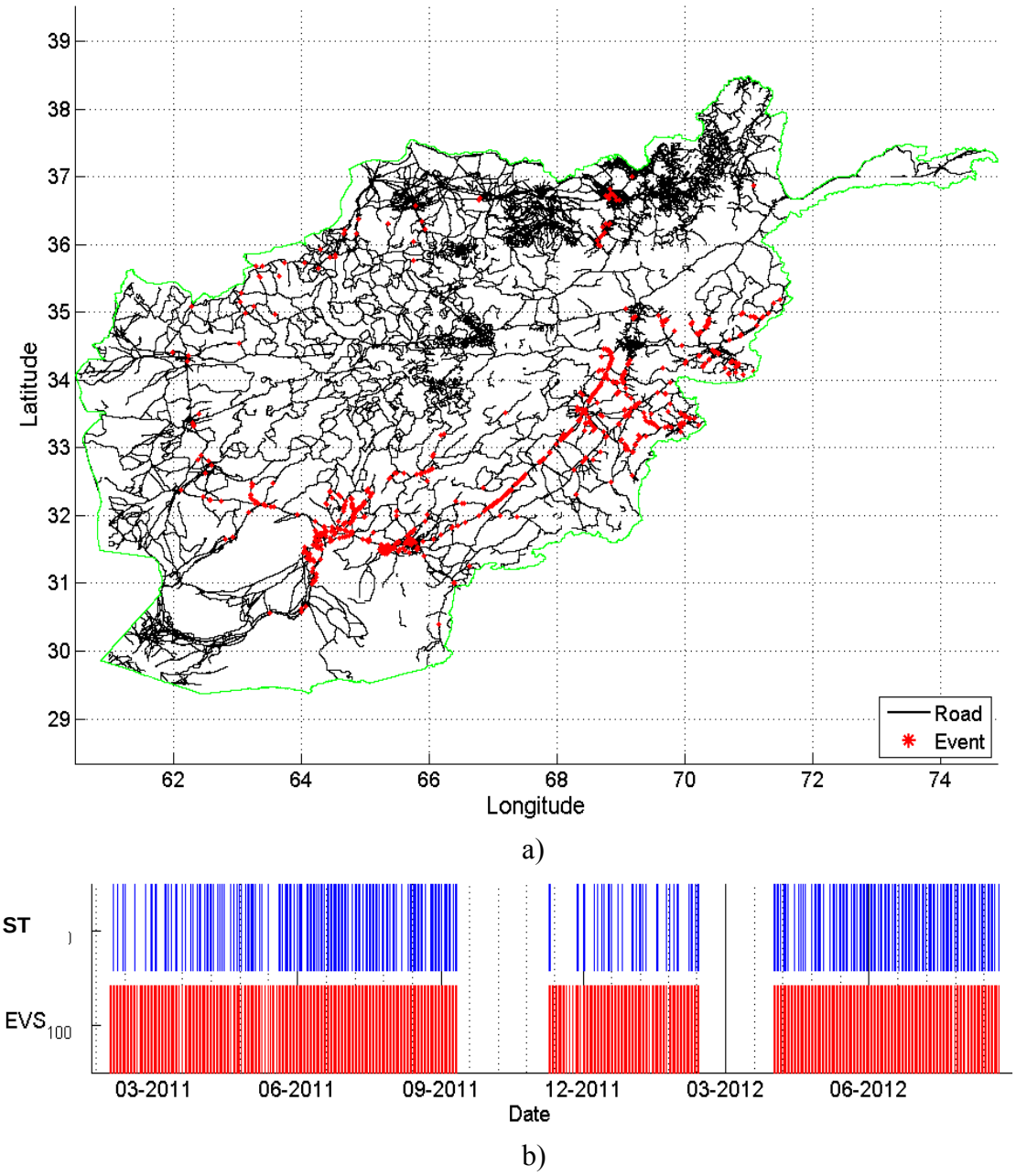



**Figure 24. Distribution of random sample of historical events, *ST*, used for analysis and parameter discovery of the MECH classification algorithm; a) geographic distribution of *ST* locations and roads; b) temporal distribution of *ST* compared to the temporal distribution of $EVS_{100}$.**

from a distance, or the location of the person reporting the event. In the case of a military patrol, a person reporting the event may be a hundred or more meters away, since patrols typically maintain 25-meter or greater spacing between vehicles. The location chosen to represent a direct fire event is similarly unclear. The reported location may be the first vehicle in the patrol, the first vehicle to come under fire, the vehicle of the person reporting the incident, or an estimated 'center' of the kill zone or ambush site. The 250-meter-radius exclusionary zone around known conflict event sites provides a buffer that attempts to mitigate these problems.

It is important to note another significant issue with the $RD_{250}$ dataset. $IED_{100}$ and $DF_{100}$ contain single classes of data: locations and dates where known conflict events occurred. The $RD_{250}$ dataset, although labeled as a single class, actually contains at least three classes:

1. Locations that are not useful for conflict events and will never be used;
2. Locations that are useful for a conflict event but have not been used yet; and
3. Locations that have already been used for a conflict event but this event occurred before or after the time period covered with this data.

Thus, when $RD_{250}$ is used as the non-event class in a two-class classification algorithm, it is likely that some misclassified non-events are actually members of (2) or (3). One impact of this issue is that the overall misclassification rate may not be a good indicator of performance. Instead, it is necessary to examine both the overall misclassification rate as well as the individual misclassification rates of each class.

The restricted conflict event sets used for learning are a reflection of an incomplete road dataset. This is evidenced by the fact that many conflict event sites are located away from known roads. Appendix A.3 illustrates this problem. The $IED_{100}$ and $DF_{100}$ datasets were chosen to ensure that the features collected from conflict event locations used for training were comparable to features collected from points known to lie along roads. The inclusionary radius of 100 meters ensures that conflict event points used for training lie sufficiently close to known roads that all features may be collected. This criterion is particularly applicable to features based on visibility and cumulative escape adjacency.

A total of 77 distinct measures derived form 20 features at various resolutions or window sizes were cataloged for each road point and conflict event location. Seven additional descriptors capture the location, time, and index of each event. The data are organized into matrices where a row represents a single location (road) or location and date (event) and a column is a distinct measure, coordinate, or index. Thus, a single row in the data consists of

$$\boldsymbol{x_i} = [\, latitude, longitude, class, index_1, index_2, date, time, f_1, f_2, \ldots, f_{77}]$$

These observations are split into Emplacement and Monitor/Control sets. Table 4 lists the features in each set. For analysis and prediction, combinations of the first seven indices are used for selection and filtering.

Classification outcomes are presented using two measures. *Percent error* is the percent of all classifications that are not correct. Cast in terms of the conventional confusion matrix, *percent error* is

**Table 4. Emplacement and Monitor/Control features**

| Emplacement Features | |
|---|---|
| Elevation | Elevation range at 50 meters |
| Slope | Roughness at 50 meters |
| Convexity | Local openness (all resolutions) |
| Texture | Distance to populated areas (all sizes) |
| Route visibility at 100 meters | |
| | |
| Monitor/Control Features | |
| Elevation range at radii > 50 meters | Visibility index (all window sizes) |
| Roughness at radii > at 50 meters | Discrete shape complexity (all window sizes) |
| Short radial (all resolutions) | Sparse VS shape complexity (all resolutions) |
| Long radial (all resolutions) | Planimetric area (all resolutions) |
| Mean radial (all resolutions) | Rugosity (all resolutions) |
| Maximum cumulative escape adjacency | Maximum route visibility at radii > 100 mtrs |
| Minimum cumulative escape adjacency | Minimum route visibility at radii > 100 meters |
| Median cumulative escape adjacency | Median route visibility at radii > 100 meters |

$$Percent\ Error = 100 * \frac{False\ Positive + False\ Negative}{All\ Positive + All\ Negative}$$

A second measure is *event classification error*, which is the percent of misclassified events. Commonly described as *precision*, it is found as

$$Event\ Classification\ Error = 100 * \frac{False\ Positive}{All\ Positive}$$

**Set Selection and Normalization**

Attackers may vary their attacks with geography and over time. Attacks may vary with geography for a number of reasons including the shape and structure of terrain, availability of critical tactical elements like overwatch sites, proximity to attacker safe

zones, and political and tribal boundaries. Over time, attacks may vary due to deployment of countermeasures, availability of war materials, experience level of the attackers, and sophistication of the target.

Geographic and temporal windows attempt to mitigate the effects of this change over distance and time. Geographic windows constrain the set of observations being used to more localized and presumably more similar terrain. The use of geographic windows is based on the assumption that nearby terrain is more likely to be used in the same way for a conflict event (similar tactics, similar geometry, similar choices made because the same people are involved), than terrain that is more distant. Temporal windows are based on a similar assumption: events that are temporally proximate are more likely to resemble each other than events that are temporally more distant, due to availability of war materials, current shared training, deployment of countermeasures, seasonal weather, and other influences that tend to vary with time.

Fixed geographic windows can be used to exclude observations that are physically distant from a specified center point. The selection criterion, geographic distance from a selected coordinate, is used to exclude observations from the set that are beyond a specified radius. Let *loc* be a location and let *radius* be a geographic distance that will be used as constraining radius. Then,

$$event\_obs = events \mid distance(loc, events) < radius \tag{34}$$

$$roads\_obs = roads \mid distance(loc, roads) < radius \tag{35}$$

In the datasets used in this research, the number of non-event observations (road points not known to be used for conflict events) is always much larger than the number

of events when using fixed geographic windows. In order to create training and test sets with balanced classes, a subsample of roads points of the same size as *event_obs* is randomly selected without replacement

$$sample = [event\_obs;\ rand\_select(roads\_obs, k)] \tag{36}$$

$$k = |event\_obs|$$

Cross-validation is used to select training and test subsets randomly from *sample*. In the following sections, three-way cross validation is used.

Fixed temporal windows can be used to exclude observations that are distant in time from a specified date. The selection criterion, timespan in days before and after the specified date, is used to exclude observations from the set that are older or newer than a selected date range. Let *date* be a calendar date that falls within the date range of available data and let $span_1$ and $span_2$ describe the timespans in days used to constrain the training and test sets, respectively. The training set includes events that occur on or before *date* up to a timespan of $span_1$ days and the test set includes events that occur up to $span_2$ days after *date*. Note that the *timespan* function returns negative values for dates that occur before *date*. Then,

$$trn\_event\_obs = events \mid -span_1 \leq timespan(date, events) \leq 0 \tag{37}$$

$$tst\_event\_obs = events \mid 0 < timespan(date, events) \leq span_2 \tag{38}$$

Because there is no geographic constraint, a subsample of roads points of the same size as *trn_event_obs* is randomly selected without replacement from all roads.

$$trn\_sample = [trn\_event\_obs;\ rand\_select(roads, k)] \tag{39}$$

$$k = |trn\_event\_obs|$$

A subsample of road points is selected similarly for the test sample.

Combined constraints produce training and tests sets that satisfy two conditions: geographic and temporal. The training set produced by these constraints contains roads and events that are within *radius* of a specified l*oc*. Training events are further constrained to be within a specified timespan before or on *date*. Let *loc* and *date* describe the geographic location and date of some event and let *radius*, $span_1$ and $span_2$ describe the geographic window radius, training timespan and test timespan. Then the training sample is found by

$$event\_obs = events \mid distance(loc, events) < radius \tag{40}$$

$$trn\_event\_obs = event\_obs \mid -span_1 \leq timespan(date, event\_obs) \leq 0 \tag{41}$$

$$roads\_obs = roads \mid distance(loc, roads) < radius \tag{42}$$

$$trn\_sample = [trn\_event\_obs;\ rand\_select(roads\_obs, k)] \tag{43}$$

$$k = \mid trn\_event\_obs \mid$$

The test set is found similarly, using $span_2$ as the timespan.

Although the use of geographic and temporal constraints produces sets of events and terrain that are probably more homogeneous, a side effect is that the training and test sets are temporally disjoint. In some parts of the year, this means that the test events will occur in a completely different season from the training events. Minimizing this seasonal difference should be considered when $span_1$ and $span_2$ are chosen. A related problem is that smaller windows tend to produce smaller data sets. For some machine learning algorithms, small training sets may be difficult to use or may produce unreliable results. Figure 25 shows the number of events found using geographic, temporal and combined

constraints. In the figure, the red crosses denote the number of events found using the 500 locations and dates in *ST*. Within each window (geographic, temporal or combined), the crosses are dithered slightly to facilitate visual assessment of the distribution of values. The blue circle denotes the mean number of events in the window. As might be expected, the number of events tends to decrease as the window size decreases. In the case of combined constraints, the smallest combinations produce few usable sets.

### *Data Normalization*

Once constraints have been applied, the resulting datasets need to be prepared for use with in machine learning techniques. The raw data must be centered and scaled so that the resulting features share a common mean at or near zero and approximately equal ranges. Three normalization methods were explored: z-score and two min-max schemes.

Z-score is probably the most widely used normalization technique. Calculated using mean for location and standard deviation for scale, the z-score for a single feature or measure *s* is found by

$$s_i' = \frac{s_i - \mu(s)}{\sigma(s)} \quad \forall\, i \;\; in\; s \tag{44}$$

The use of z-score is not optimal in every case. The normalized output for different features may have different numeric ranges. Also, z-score normalization is sensitive to outliers due to the use of the feature mean and standard deviation. Z-score implicitly assumes that the distribution of *s* is Gaussian. For features with distributions other than Gaussian, mean and standard deviation may not be reasonable for location and scale. Also, the original distribution is not necessarily preserved.

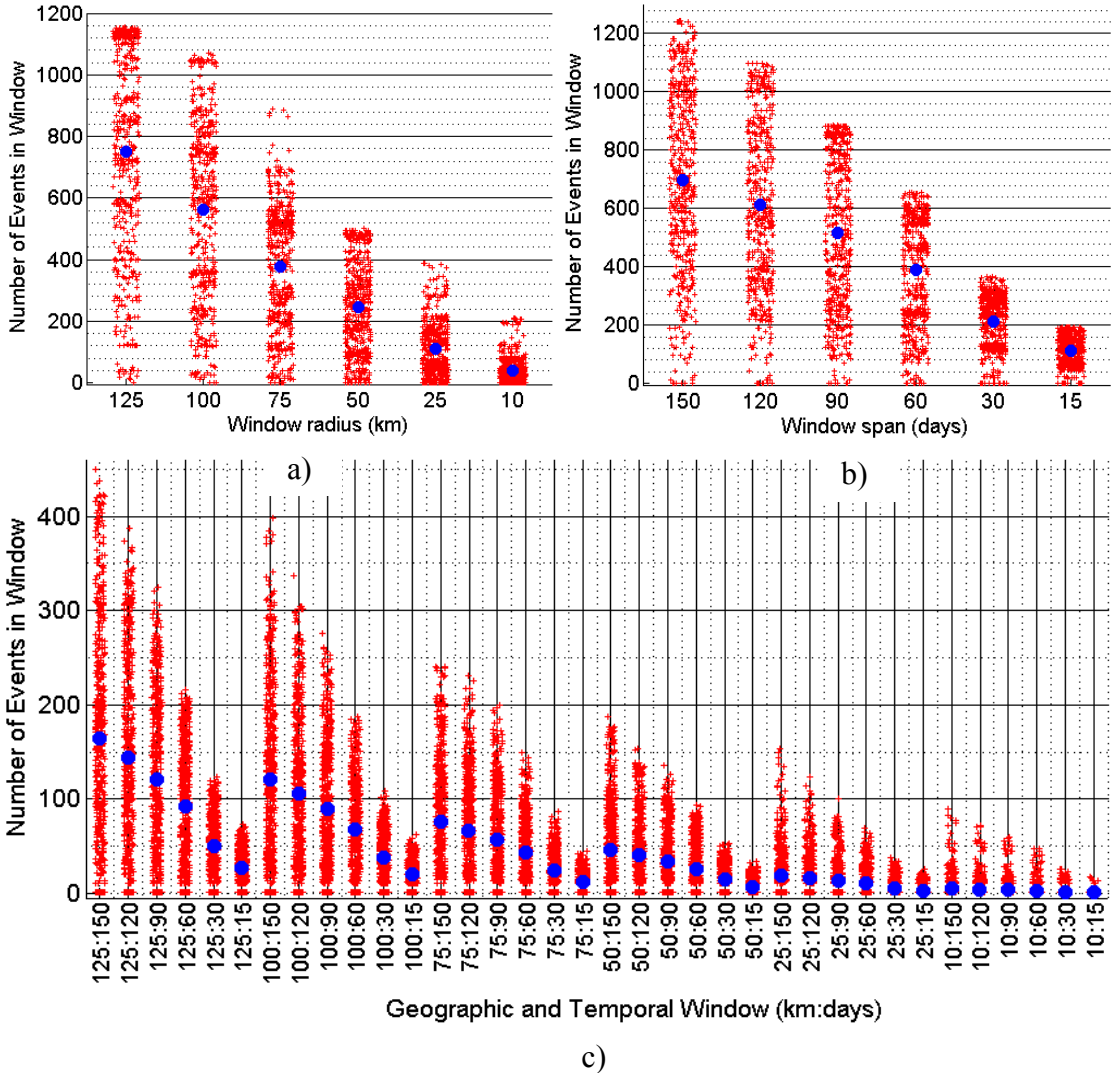



**Figure 25. Number of events found using various constraints; a) geographic; b) temporal; c) combined geographic and temporal.**

Min-max is another widely used normalization technique. In min-max, the feature mean provides the location information while the difference between maximum and minimum feature values provides the scale. Min-max normalization for a single feature *s* is found by

$$s_i' = \frac{s_i - \mu(s)}{max(s) - min(s)} \quad \forall\ i\ \ in\ s \tag{45}$$

As with z-score, min-max normalization is very sensitive to outliers. However, min-max scaling does preserve a scaled and center version of the original distribution.

Both z-score and min-max are sensitive to outliers due to their reliance on the mean and standard deviation. An alternate version of min-max, referred to here as 3-sigma avoids this problem by using the median of the feature as its location, instead of the mean. Standard deviation is calculated using the subset of $s$ between the first and ninth deciles (10th and 90th percentiles). The standard deviation and median are used to set the min and max feature values to +/- 3*standard deviation. This ensures that most of the data (80%) lie within the range [0, 1] with outliers assigned values above and below these bounds. In this normalization technique, the denominator $(median + 3 * \sigma(feature)) - (median - 3 * \sigma(feature))$ reduces to $6 * \sigma(s)$. Therefore, 3-sigma normalization is found by

$$s_i' = \frac{s_i - median(s)}{6 * \sigma(s'')} \quad \forall\ i\ \ in\ s, \quad s'' = \ decile_1 \leq s \leq decile_9 \tag{46}$$

A criticism of 3-sigma is that the use of standard deviation still causes an implicit assumption of normally distributed data. Also, as shown in Figures 12 and 17, exclusion of the first or tenth decile may remove a significant fraction of events from inclusion in the principal scaling parameter, *s''*, especially when all data is used to develop location and scale parameters. Finally, location and scale parameters may exhibit greater variance when small samples are further reduced by discarding the first and tenth deciles.

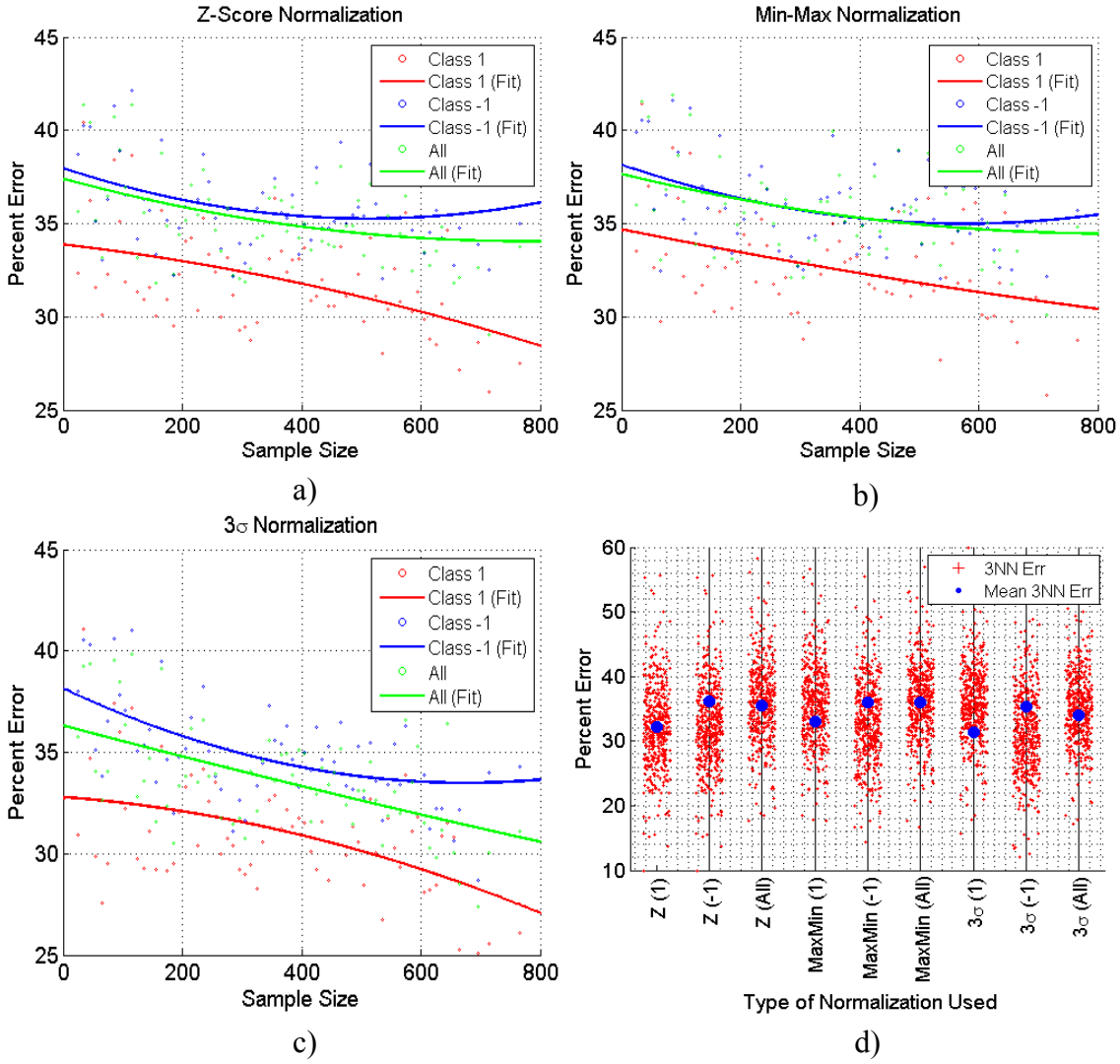


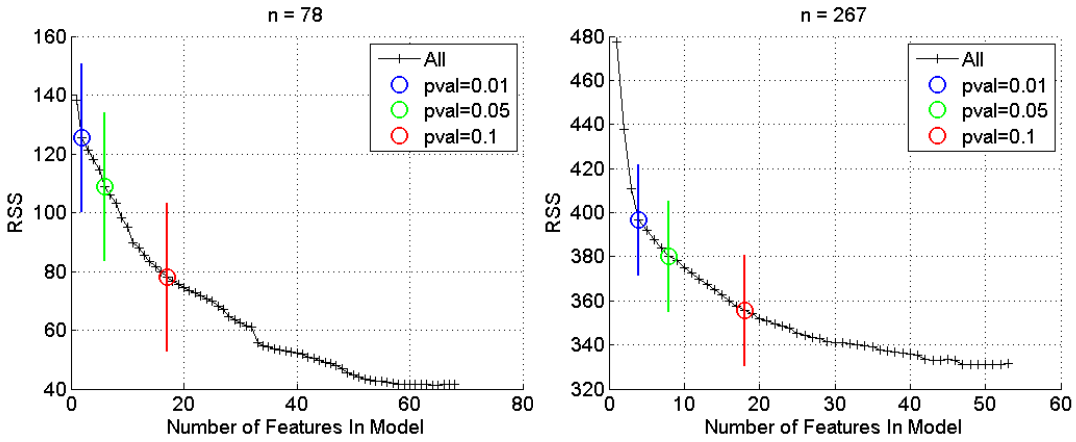

**Figure 26. Comparison of normalization schemes; a) z-score; b) min=max; c) 3-sigma; d) comparison of means and distribution across all schemes using 3NN.**

Normalization schemes typically operate on all data, in an unsupervised manner. However, as noted in [134], local scaling based on the selection of observations from a 'neighborhood' can enhance class separation and improve classification performance. A procedure used in applications as varied as detecting audio similarity [135] and clinical

trials of new drugs [120], selection of a homogenous data sample for determination of the location and scale parameters is an effective way to improve classification performance. For this research we examine the impact of using individual classes to calculate location and scale parameters.

Figure 26 examines the impact of normalization on classification accuracy using the 3-nearest neighbors (kNN, k=3) classification algorithm on data that is constrained both geographically and temporally. Three different sample subsets were used to produce the location and scale parameters that were then used to normalize the entire dataset: Class 1 (location and scale from IED events), Class 2 (location and scale from roads) and all data (location and scale from all data). While the average classification error varies between 31-36% across all normalization schemes, Figure 26.d shows that the best performance in every normalization scheme is found when Class 1 is used to produce the location and scale parameters. Figure 26.a-c reinforces this idea, showing that best classification accuracy is found when the location and scale factors are calculated using the event data (Class 1).

Although 3-sigma slightly outperforms z-score, for the remainder of this dissertation we use z-score for normalization. This choice is driven by two factors. First, we accept that sample sizes are likely to be small when combined geographic and temporal filters are used. This is potentially a problem with 3-sigma. Also, the small performance gain, less than 1%, fails to compel a change from the more widely known and commonly used *z*-score. Class 1 (events, either IED or DF) will be used to produce the location and scale factors.




a) b)

**Figure 27. Number of features selected by stepwise feature selection using various *penter* thresholds; a) a smaller dataset with 78 observations; b) a larger dataset with 267 observations.**

### Feature Selection and Dimensionality Reduction

In order to predict accurately, it is probably important to select features that are relevant and useful for supporting classification decisions. In this research we examine the utility of two different mechanisms for achieving this goal, stepwise feature selection and principal component analysis. Stepwise feature selection produces an optimal or near-optimal subset of features that increases classification accuracy. The subset produced is typically smaller that the set of all features. Principal component analysis converts a set of possibly correlated features or measures into a set of linearly uncorrelated variables organized by variance. The subset of PCA-produced variables accounting for most of the data variance is typically smaller than the set of all features.

*Stepwise Feature Selection*

Stepwise feature selection (SF) is a supervised method for selecting a subset of features from a larger dataset [136]. With a goal of finding an optimal or near-optimal set of features to support accurate classification of data, SF uses linear regression to evaluate combinations of features in order to increase classification accuracy. A result of stepwise feature selection is a model that provides 'big picture' view of feature relevance and excludes elements that do not contribute significantly [81].

As used in this research, stepwise feature selection begins with an empty model with no features included. Individual features are added to, and potentially removed from, the model in sequential steps based on the significance, or *p*-value of the two-tailed Student's t-test,

$$SE\left(\hat{\beta}\right) = \frac{\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i-\hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \qquad ttest = \frac{\hat{\beta}}{SE\left(\hat{\beta}\right)}, \qquad pvalue = 2 * cdf(|ttest|) \qquad (46)$$

, where $x$ is the same, single feature across all observations and $\bar{x}$ is the feature's mean. $Y$ is the real class and $\hat{y}$ is the predicted class. $SE\left(\hat{\beta}\right)$ is the standard error of the feature's estimated regression coefficient. CDF is the cumulative distribution function of the Student's t-distribution. A small *p*-value rejects the hypothesis that the feature would have a coefficient of zero if added to the model. At each step, using Matlab's *stepwisefit* function [137], all features not in the model are evaluated. The feature with the smallest *p*-value is added to the model provided it is greater than a threshold. This threshold is governed by the *penter* parameter which is set to a default of 0.05 in this research. After the feature has been added, all features currently in the model are reevaluated and any

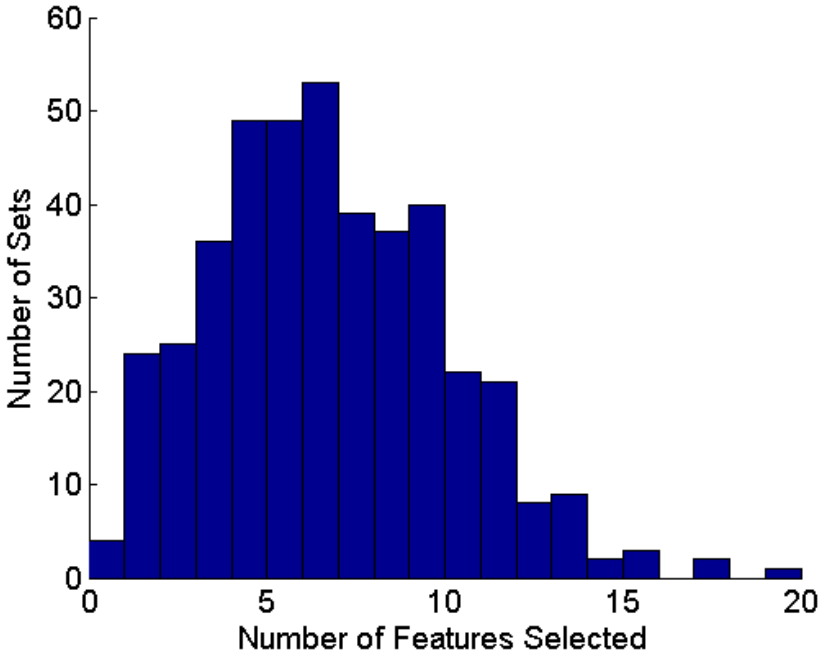



**Figure 28. A histogram shows the distribution of the number of features selected from geotemporally constrained training sets using stepwise feature selection with *penter* = 0.05 and *premove* = 0.10.**

feature with a *p*-value greater than *premove*, set to a default of 0.10 in this research, are removed. SF continues until no remaining features produce a *p*-value less than *penter* and no features in the model have a *p*-value greater than *premove*.

Figure 27 examines the impact of the value of the *penter* parameter on the size of the final model. Three different *penter* values were used in SF to select features. The *premove* parameter was set to 0.1 for all runs. In the figure, smaller values for *penter* caused the selection of much smaller models. However, the selection of very small *penter* values selected models with few features that did a poor job of classification (shown as high RSS in the figure).

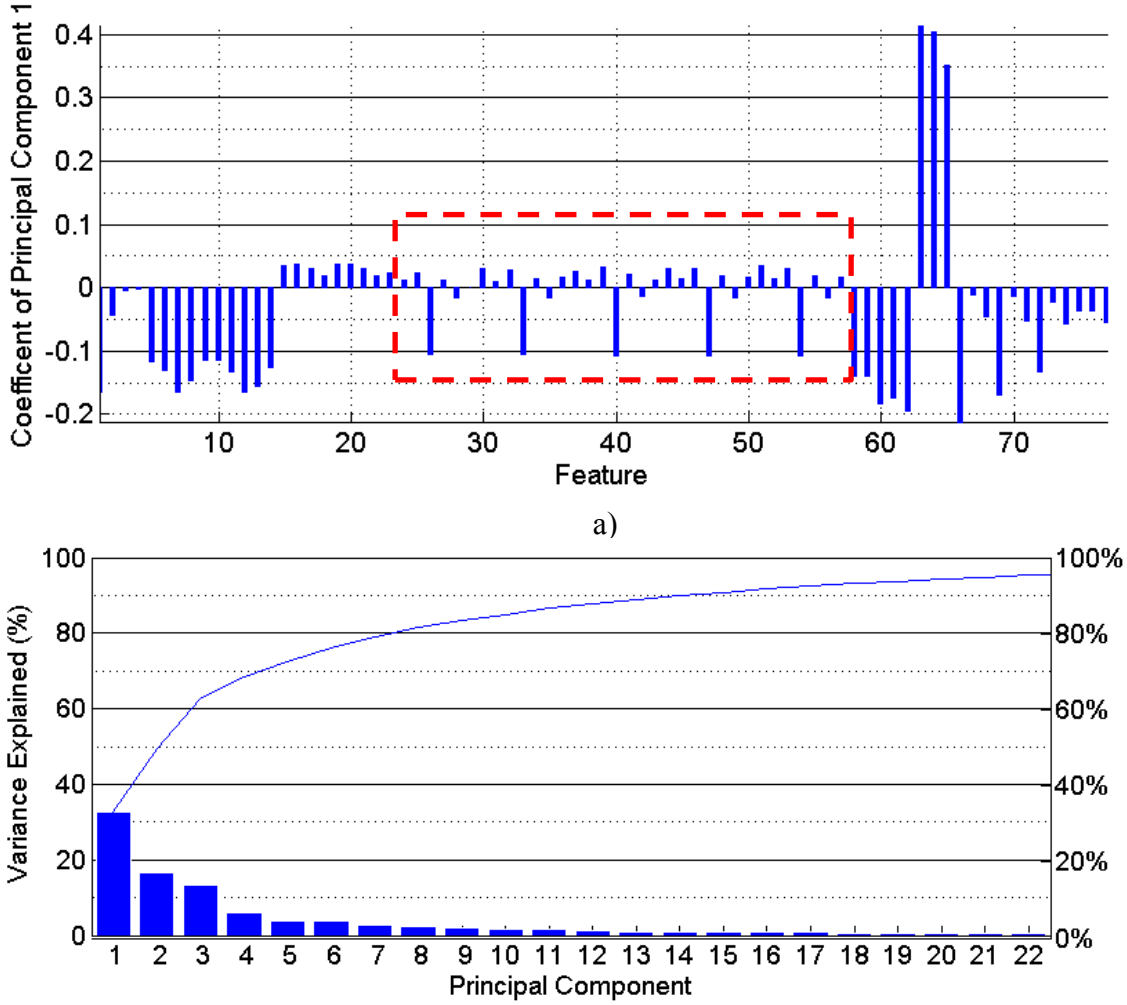



**Figure 29. Principal component analysis; a) PCA loadings (coefficients) for a first principal component for a randomly selected IED event; b) Pareto graph of the variance explained by principal components.**

Figure 28 examines the number of features selected by FS from training sets using locations and dates from *ST*. For each event in *ST*, a training set was chosen using a geographic constraint of 100 km and a temporal constraint of 120 days. SF was used on each training set to select features with a *penter* parameter of 0.05 and a *premove* parameter of 0.1. The number of features selected across all events in *ST* tended to be

fairly small. From a total of 77 features, FS tended to select between 2 and 12 features for inclusion in the model. In no case was more than 20 features selected.

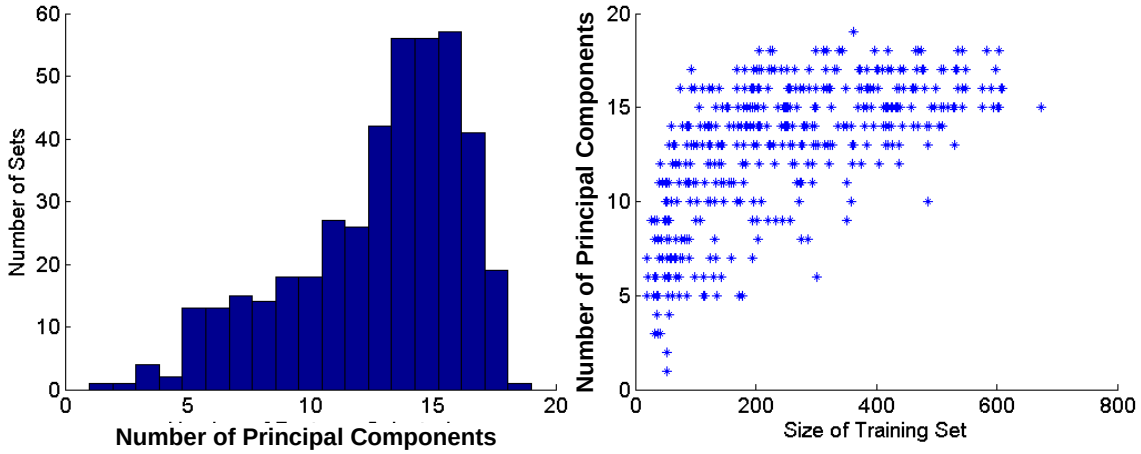
SF was selected for use in this research because it has been a common and useful choice in feature selection and dimensionality reduction for years [113], [138], [139]. However, there are potential problems associated with its use including underestimation of *p*-values and exacerbation of collinearity problems. A less objective but more damning accusation is that SF encourages thoughtless or blind application of an algorithmic approach that might not be suited to the data or that might be outperformed by an expert familiar with the problem [117].

### *Principal Component Analysis*

Principal component analysis (PCA) is an unsupervised method of converting a set of possibly correlated features into set of linearly uncorrelated variables. The orthogonal transformation used by PCA produces variables called principal components that are guaranteed to be independent when the original features are normally distributed and independent [140], [141], [142]. A simple, non-parametric method of producing relevant, independent variables, PCA provides a lower-dimensional representation of a higher-dimensional dataset where the first principal component accounts for as much of the variability in the data as possible. Each successive principal component accounts for as much of the remaining variability as possible.

In order to produce the principal components, centered data *x* with *m* features is used to calculate the *m* x *m* covariance matrix Σ,

$$\Sigma_{i,j} = covariance\left(x_{:,i}, x_{:,j}\right), \quad for \quad i,j = 1, \dots, m \tag{47}$$




a) b)

**Figure 30. Number of principal components found in geotemporally constrained training sets using PCA; a) histogram showing distribution of the number of principal components required to account for 90% of variance in the dataset; b) The size of the training set impacts the number of features selected.**

, where $x_{:,i}$ and $x_{:,j}$ are individual features (columns) across all observations in *x*. The covariance matrix is used to determine the eigenvalues and eigenvectors. The eigenvectors are sorted in order of decreasing eigenvalue, an ordering that produces a matrix of principal component coefficients or loadings. The loadings and *x* are multiplied to produce the score or final data set.

Figure 29 examines the results of using PCA on a single training set. The magnitude of the coefficients selected by PCA for the first principal component can be seen in Figure 29.a. In the figure, PCA chooses coefficients ranging from -0.21 to 0.41. (An interesting aside is the pattern seen inside the dashed red box. This pattern is an artifact of the order in which features were added to the dataset. Seven features collected at the same resolution were added as a group and a total of five different resolutions were used. PCA assigned very similar coefficients to each of the five groups of seven

features.) Figure 29.b examines the amount of variance explained by the principal components. In the figure, the first principal component accounts for over 30% of all variance in the dataset. Successive principal components account for increasingly smaller amounts of variance. For this particular dataset, the first 14 principal components account for a total of approximately 90% of the variance in the data

Figure 30 examines the number of principal components found PCA in training sets using locations and dates from *ST*. For each event in *ST*, a training set was chosen using a geographic constraint of 100 km and a temporal constraint of 120 days. In Figure 30.a, the number of principal components required to account for 90% of variance in the dataset tended to range between 6 and 18. In all cases, less than 20 principal components were required to account for 90% of the variance in the dataset. Figure 30.b notes the relationship between the size of the training set and the number of features selected. In general, as the size of the training set increased, the number of features selected also increases.

PCA makes a number of assumptions that impact its performance and the utility of the loadings produced. First, PCA assumes that interesting differences or dynamics are indicated by large variances and that small variances represent noise or uninteresting differences. This may not be true in many cases. PCA also assumes that mean and variance are sufficient statistics to describe or capture a probability distribution. In fact, these statistics only describe a Gaussian distribution. If the data have some other distribution, PCA may fail to capture important patterns. This effect is probably greater in smaller datasets. A final assumption mentioned here is that PCA creates variables

from weighted linear combinations of features. In many cases, significant relationships or patterns in the data may have nonlinear characteristics that PCA masks or erases.

## Classification Using Machine Learning Algorithms

Once a constrained set of data has been identified and, if desired, features have been selected, machine learning is used to develop a classifier. The classifier is trained on a portion of the data and tested on the remainder. The outcome of the test is used to estimate the learner's ability to used limited training data to build a general model that accurately predicts group membership (event vs. non-event) in previously unseen test data. In this research we examine the usefulness of three machines learning algorithms:

- support vector machines using linear and radial basis function kernels
- linear and quadratic discriminant analysis
- k-nearest neighbors with a Euclidean distance function

### *Support Vector Machines*

A support vector machine (SVM) is a non-parametric, supervised learning method used in this research to perform binary classification. As initially proposed by Vapnik and then revised by Cortes and Vapnik [91], [92], SVM maps training data as points in space and uses an optimal hyperplane to separate by the largest possible margin the classes found in the data. New observations are mapped into the same space and assigned a class based on their location with respect to the hyperplane.

SVM is implemented in this research as a two-norm soft margin classifier [81]. The optimal hyperplane separating the data is defined as

$$y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \qquad i = 1, \dots, n \tag{48}$$

, where $y_i \in \{-1,1\}$ is the class label; $x_i$ is a single observation from the training data; $\beta$ is a weight vector; and $\xi_i \geq 0$ is a slack variable that compensates for non-separable data. This optimal hyperplane is found by minimizing $\beta$,

$$\min \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n}\xi_i \quad s.t.\begin{cases} y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, \\ \xi_i \geq 0, \quad \sum \xi_i \leq constant \end{cases} \quad i = 1, \dots, n \tag{49}$$

Two kernels are used in this research, linear and Gaussian radial basis function (RBF). The linear kernel is a simple dot product, often chosen when the data is believed to be linearly separable or as a basis of comparison with other kernels,

$$K(x_i, x_j) = x_i \cdot x_j \qquad i, j = 1, \dots, n \tag{50}$$

The Gaussian radial basis function is another commonly used kernel function

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \qquad i, j = 1, \dots, n \tag{51}$$

In this function, $\|x_i - x_j\|^2$ is the squared Euclidean distance between $x_i$ and $x_j$. The $\sigma$ parameter, also referred to as the shape or spread parameter, controls the shape of the RBF kernel. Smaller values of σ correspond to a flatter or wider basis function while larger values produce a steeper or narrower basis function [105][106]. The shape parameter is associated with the fit of the model to the training set. If σ is set to zero, every training instance becomes a support vector and over-fitting occurs. The training set is perfectly predicted but the test set may have high classification error. On the other hand, if σ is set to ∞, all training and test instances are collapsed into a single class.

*Discriminant Analysis*

Discriminant analysis is a parametric, supervised learning method used in this research to perform binary classification. Discriminant analysis finds a linear combination of features that separates two classes by explicitly modeling differences between the classes of data. In this research, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are used.

In LDA, the decision boundary is determined by the covariance of the class distributions and the positions of the class centroids. As described in [81], the linear discriminant function for class *k* is

$$\delta_k(x) = x^T\hat{\Sigma}^{-1}\hat{\mu}_k - \frac{1}{2}\hat{\mu}_k^T\hat{\Sigma}^{-1}\hat{\mu}_k + log\ \hat{\pi}_k \tag{52}$$

The parameters of the Gaussian distribution for class *k* are estimated as

$$\hat{\pi}_k = n_k/n \quad n_k = number\ of\ observations\ of\ Class\ k\ in\ n, \tag{53}$$

$$\hat{\mu}_k = \sum_{Class=k} x_i/n_k \quad , \tag{54}$$

$$\hat{\Sigma} = \sum_{k=1}^{K}\sum_{Class=k}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T/(n-K) \quad \forall i \in x \tag{55}$$

Note that the linear discriminant function shares a common covariance matrix for all classes (Equation 55). Given Equation 52, then the decision boundary with two classes, *K*=[1,2], can be described as

$$\delta_{k=1}(x) = \delta_{k=2}(x) \tag{56}$$

and the LDA rule classifies to class 1 when

$$x^T\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) > \frac{1}{2}\hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_1 - \frac{1}{2}\hat{\mu}_2^T\hat{\Sigma}^{-1}\hat{\mu}_2 - log\ \hat{\pi}_1 + log\ \hat{\pi}_2 \tag{57}$$

In QDA, the decision boundary is determined by the individual covariance of each class and positions of class centroids. Noting that the covariance matrix is different for each class (Equation 58-59), the quadratic discriminant function for class *k* is

$$\delta_k(x) = -\frac{1}{2} log|\Sigma_k| - \frac{1}{2}(x - \hat{\mu}_k)^T \hat{\Sigma}_k^{-1}(x - \hat{\mu}_k) + log\ \hat{\pi}_k \tag{58}$$

$$\hat{\Sigma}_k = \sum_{Class=k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T/(n - K) \quad \forall i \in x \tag{59}$$

The decision boundary is described as in Equation 56.

*k-Nearest Neighbors*

The k-nearest neighbor algorithm (kNN) is a non-parametric method of classifying a sample or observation with an unknown classification. A new observation is assigned to the same class as the majority of its nearest neighbors. Perhaps the simplest of all classification algorithms, kNN makes no assumptions about prior or posterior distributions and requires only the selection of a distance metric and a single hyperparameter, *k*, which fixes the number of nearest neighbors used to determine class [103].

In kNN, the distance metric describes a measure of similarity or difference between two samples. A commonly used, intuitive metric is Euclidean distance, where the difference or distance between two samples is length of the shortest line that would connect them. It is calculated using the Pythagorean formula,

$$distance(x_i, x_j) = \sqrt{\sum_{k=1}^{m} (x_{i,k} - x_{j,k})^2} \tag{60}$$

, where $x$ is a matrix containing $n$ samples, each with $m$ features and $x_{i,k}$ is the $k$th feature of the $i$th sample in $x$. Ghosh notes in [143] that Euclidean distance performed best in a majority of datasets tested. Euclidean distance is the metric used in this research.

A second important choice is the order of $k$ the number of nearest neighbors used to assign a class to a new observation or sample. Typical values of $k$ range from a minimum of 1 to maximums of $n^{1/2}$ and $n^{5/8}$ [144], and $2n^{1/2}$ [143]. All of these sources note that an empirically selected value of $k$ somewhere between these extremes is likely to have the best classification accuracy.

kNN is known to have two significant problems. First, at high dimensionality the distance between pairs of points tends to converge. This problem is exacerbated when features are highly correlated or noisy. Second, kNN performs best when the number of samples is large. When the training set is small, or when one class in the training set is significantly smaller than the other, the classification accuracy of kNN decreases.

**Experimental Overview**

In the following sections, we step though the MECH Classification algorithm sequentially, building and supporting the case for the final solution that incorporates a range of dynamic elements. In the initial analysis of the impact of geographic, temporal and geotemporal constraints, an experiment is constructed that uses common, shared default parameters. This ensures that any differences noted can be attributed to changes in constraints and not to parameter selection or algorithmic optimizations. The settings for the default experiment are:

- Geographic windows (in km): 125, 100, 75, 50, 25, 10

- Temporal training windows (in days): 150, 120, 90, 60, 30, 15
- Temporal test window (in days): 60
- PCA: all principal components are used
- Stepwise Feature Selection: all weighted features are used
- SVM with linear kernel: box constraint=1
- SVM with RBF kernel: box constraint=1, RBF shape parameter=1
- kNN: $k = 1, 3$

A variety of optimizations are explored including:

- Dimensionality reduction
    - PCA: using cumulative variance to determine the number of principal components
    - STP: using p-values to constrain the size of the final feature set
- Parameter estimation
    - SVM: dynamic selection of box constraint and RBF shape parameter
    - kNN: dynamic selection of $k$
- Set selection using dynamic geographic and temporal constraints
- Unbalanced classes
- Feature selection using experts and the MECH Model
- Combinations of classifiers

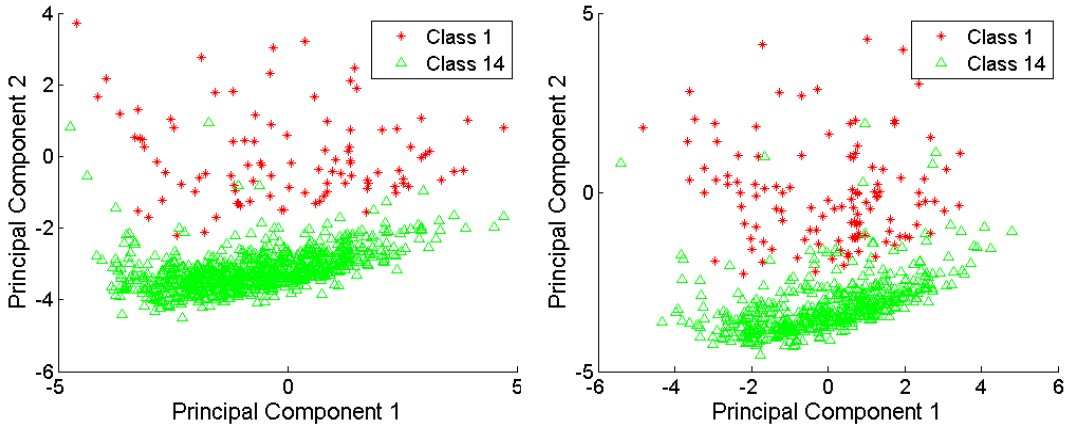
## The Impact of Fixed Geographic and Temporal Constraints

Selection of the training set is one of the most important steps in machine learning. A well-chosen training set coupled with an appropriate learning algorithm will produce useful classification rules that accurately assign new data to the correct class. For MECH, two factors constrain the range of events that should be included in the training set. These constraints are geographic, where the nature of terrain drives the choice of tactics and or attack configuration, and temporal, where seasonal variations, material availability and human factors cause change over time.

### *Geographic Constraints*

Geography and associated human factors have an undeniable impact on conflict event tactics. Tactics suitable for use on flat ground are ill-suited for use in mountainous terrain. For example, ambush hide sites near the road may not be available flat terrain while control sites with good visibility may not be available in mountainous terrain. Thus, the morphographic structure of terrain forces or constrains choices related to conflict event siting and execution. Since terrain tends to change slowly, intuition suggests that locations that are closer to each other are likely to be geographically similar and to offer similar support for certain tactics or attack configurations.

Geography also has social and cultural influences. The population in rural Afghanistan in often organized into groups or clans. These groups share training, tactics and weaponry. They work under a single leader or leadership group and may have common or shared access to war materials like explosives. The boundaries of territories under the control of these groups are often formed by rivers, mountains, and other




a) b)

**Figure 31. Conflict events in different types of terrain produce distinct patterns; a) IED events; b) direct fire events.**

dominating geographic elements. Thus, for a geographic region, bounded by natural features and dominated by a particular group, it seems possible that the type and configuration of conflict events will be more homogeneous than events from distant regions.

Figure 31 shows that conflict events that occurred in different terrain are described by distinct patterns. In the figure, Class 1 is an Iwahashi and Pike terrain classification [22] that describes steeper, very convex steep ground. Class 14 describes terrain that is flatter, and less convex. The figures were produced by first normalizing and performing PCA on all 77 features of the conflict events that occurred in Class 1 terrain. The normalization factors and PCA loadings were then used to transform the events that occurred in Class 14 terrain. Each event was plotted using the first two principal components.

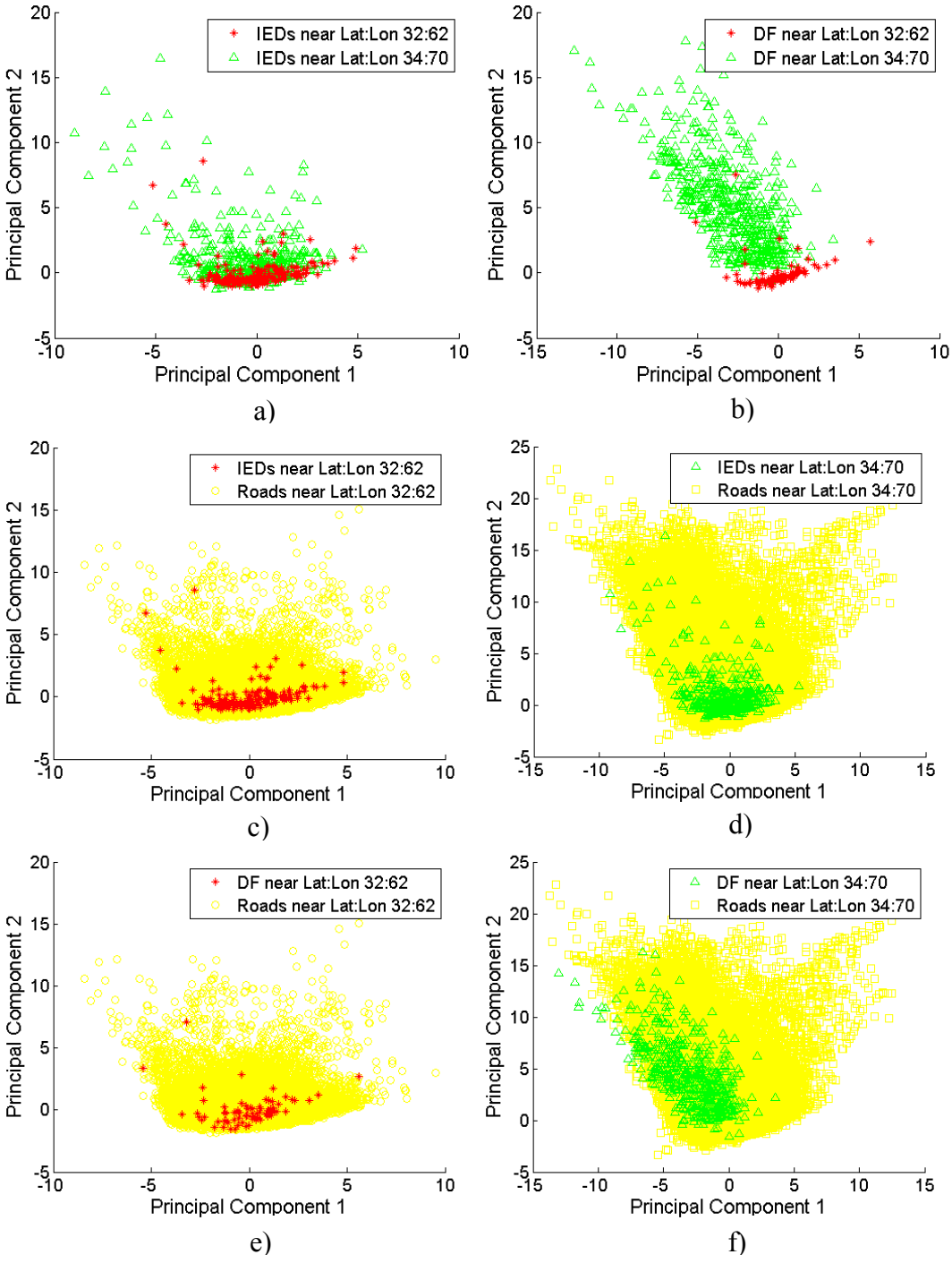



**Figure 32. Conflict events from distant locations produce distinct patterns; a) IED events; b) direct fire events; c) IED events vs. roads near 32:62; d) IED events vs. roads near 34:70; e) DF events vs. roads near 32:62; f) DF events vs. roads near 34:70.**

In the figure, it is clear that events occurring in flatter terrain are easily distinguished from events in rougher, steeper terrain. It is also interesting to note that the data for events in steeper terrain tend to be more dispersed while the data for events in rougher, steeper terrain tend to be more clustered.

Figure 32 shows that events geographically distant from each other are also described by different patterns. In the figure, IED and DF events near point [32.5N, 062.5E] are normalized and then transformed using PCA. The resulting normalization factors and PCA loadings are used to transform IED and DF events that occurred near [34.5N 070.5E]. Each event was plotted using the first two principal components. The two center points are approximately 775 km apart. In Figure 32.a, IED events that occurred near [32.5N, 062.5E] form a relatively tight cluster. The IED events that occurred near [34.5N 070.5E] overlap with the first cluster but show much greater dispersion. The DF events in Figure 32.b show a distribution like IED events but have significantly less overlap.

One key item to note about Figures 31 and 32 is that the points shown, while described as events, are actually points on the ground where an event occurred. Thus, both figures are open to alternate interpretation. It could be argued that Figure 31 simply shows that different types of terrain are described differently and that Figure 32 shows that two distant terrain points tend to be different. However, as shown in Figures 32.c-d, the cluster of IED events occupies an area much smaller than roads in the same area, indicating that IED events occupy a specific and constrained subset of roads in a given area. This supports the notion that site selection is occurring and that only a certain

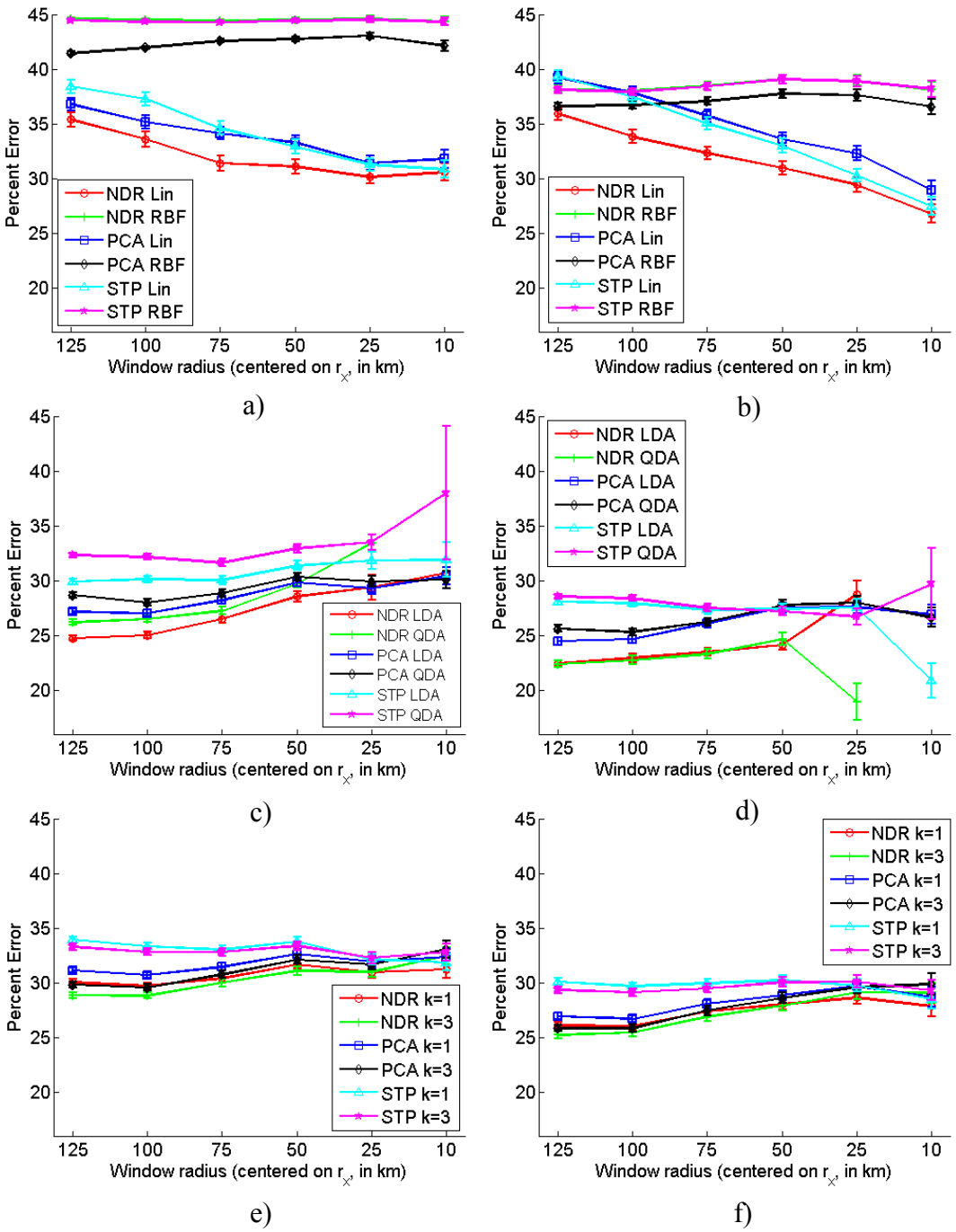



**Figure 33. The impact of window size on classification accuracy for IED and direct fire events using geographic windows ranging from 10-125 km; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

subset of relatively homogeneous locations are useful for events within a given area. The results are similar for DF events, as shown in Figures 32.e-f.

Figure 33 examines the impact of geographic radius on classification accuracy for IED and DF events. For each event in *ST* and for various radii, when the total number of IED or DF events within the window exceeds 10, three dimensionality reduction schemes, PCA, STP and NDR are performed on normalized data and then combined with SVM (using linear and RBF kernels), discriminant analysis (linear and quadratic) and kNN (1-NN and 3-NN) to estimate the classification error, which is calculated as the mean of the percent error. Three way cross-validation is used and error bars indicate a 95% confidence interval. (The lines connecting data points are provided to increase readability of the figure.)

Three observations emerge from visual inspection of the graphs in Figure 33. First, window radius seems to have little impact on classification accuracy using kNN and discriminant analysis. So, larger training sets collected over a larger geographic area do not necessarily provide better classification. The outlier is SVM using a linear kernel, where smaller radii produce better classification accuracy. Second, DF classification accuracy seems to be consistently higher than IED classification accuracy. The reasons for this are unclear but anecdotal information suggests that direct fire attack sites may be selected to allow intersecting lines of weapons fire from multiple locations. This may constrain the range of sites that are useful for this type of event. Finally, using only geographic filters, the average classification error approaches 30% for most

combinations of classification algorithm and dimensionality reduction method. This approximate error holds across both event types.

The accuracies of discriminant analysis in Figure 33.c and 33.d are misleading because they are skewed by sample size. Only larger training sets result in a classification solution because discriminant analysis fails using smaller training sets. This is due to the fact that classification using DA involves a matrix inversion requiring that the matrix being inverted have full rank. For this to occur, the number of linearly independent columns must be greater than or equal to the number of observations. If this requirement is not met, then the discriminant analysis covariance estimates will not have full rank and, thus, cannot be inverted (without resorting to use of a pseudo-inverse). In the experiment that produced the graphs in Figure 33, discriminant analysis is only used when this full rank criteria is met. For this reason, in cases where discriminant analysis seems to classify with much more accuracy than SVM or kNN, it is important to note that discriminant analysis results exclude all cases where the number of training samples is smaller than the number of features or variables.

Figure 34 examines the impact of sample size on classification accuracy across all geographic window sizes. The figure was generated by plotting the number of events available for training against the classification error obtained using three classification methods: SVM, DA and kNN, and three dimensionality reduction mechanisms: no reduction, PCA and stepwise feature selection. In the figure, sample size and variance are inversely related with smaller sets displaying much higher variance. For DA and kNN, there is a tendency to converge on an average classification error of approximately

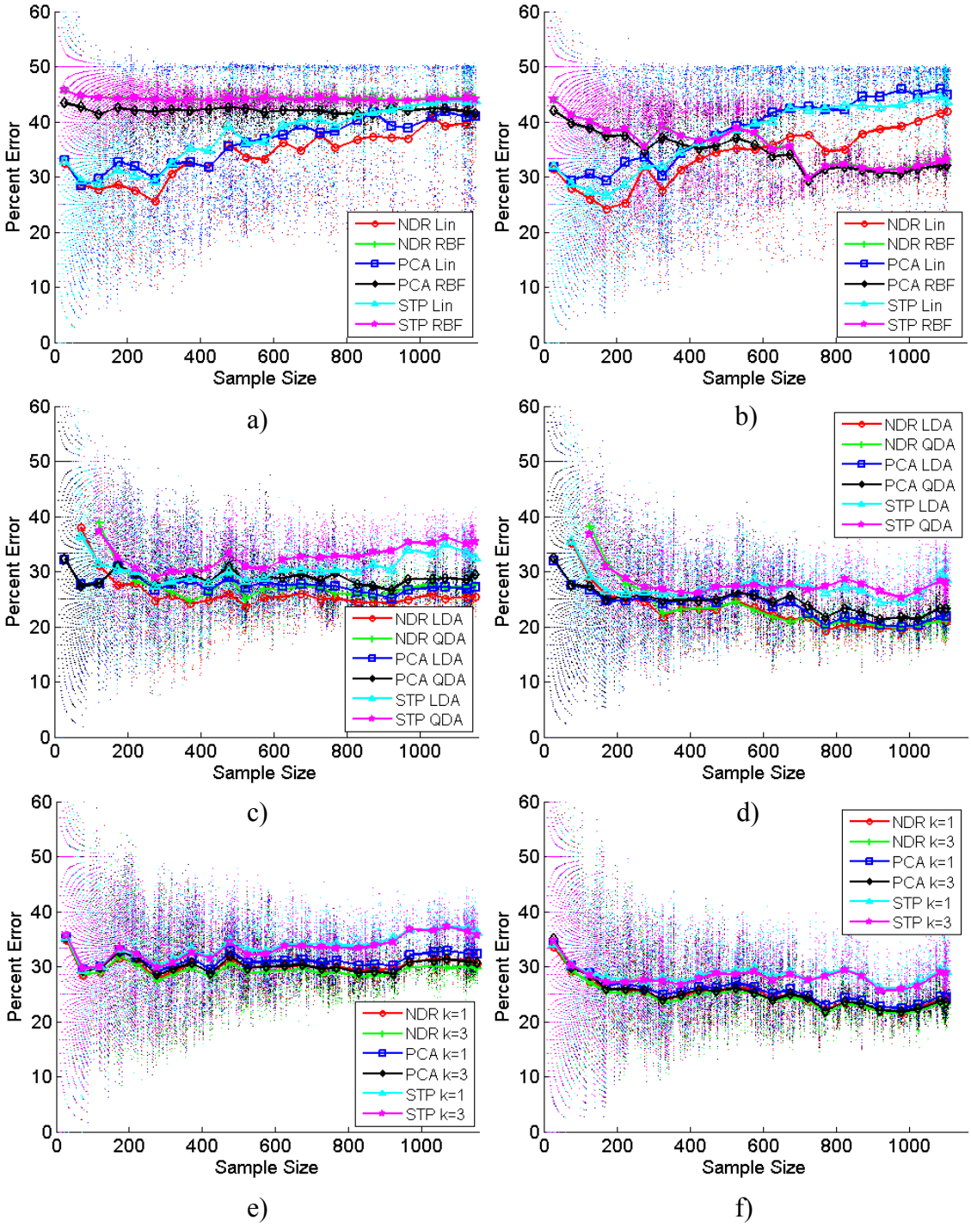



**Figure 34. The impact of sample size on classification accuracy under geographic constraints; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

30%. In the case of SVM, which appears to underperform the other classification methods, the variance remains high across all sample sizes. As before, DA classification accuracy data is only available for datasets where the sample size is relatively large. Interestingly, for all classification methods, dimensionality reduction schemes show similar performance, especially with smaller samples.

Another way to examine the impact of sample size on classification accuracy is event density. In Figure 35, event density is calculated as the number of events per kilometer of roads in the window. Unlike sample size, event density appears to have an impact on classification accuracy with some improvement seen at higher densities. Unfortunately, the number of higher density datasets is small. Interestingly, across all combinations of learning algorithm and dimensionality reduction, average classification error for IED events first increases as sample density increases to approximately 0.5, before beginning to decrease. This is not true for DF events. As previously noted, sample size and dimensionality reduction scheme skews DA performance.

Analysis of geographic constraints provides interesting insights into historical event data. The most salient insight is probably that sample size and window radius appear to have little effect on classification error, except when using SVM. On one hand, this is a positive outcome because classification based on training with smaller sets and over smaller windows is likely to be as accurate as results from much larger sets. Thus, small datasets are probably sufficient to train the classifier. On the other hand, the lack of degradation of classification accuracy over distance may indicate that the effect of geography on tactics is less important or measurable than other factors that do not vary

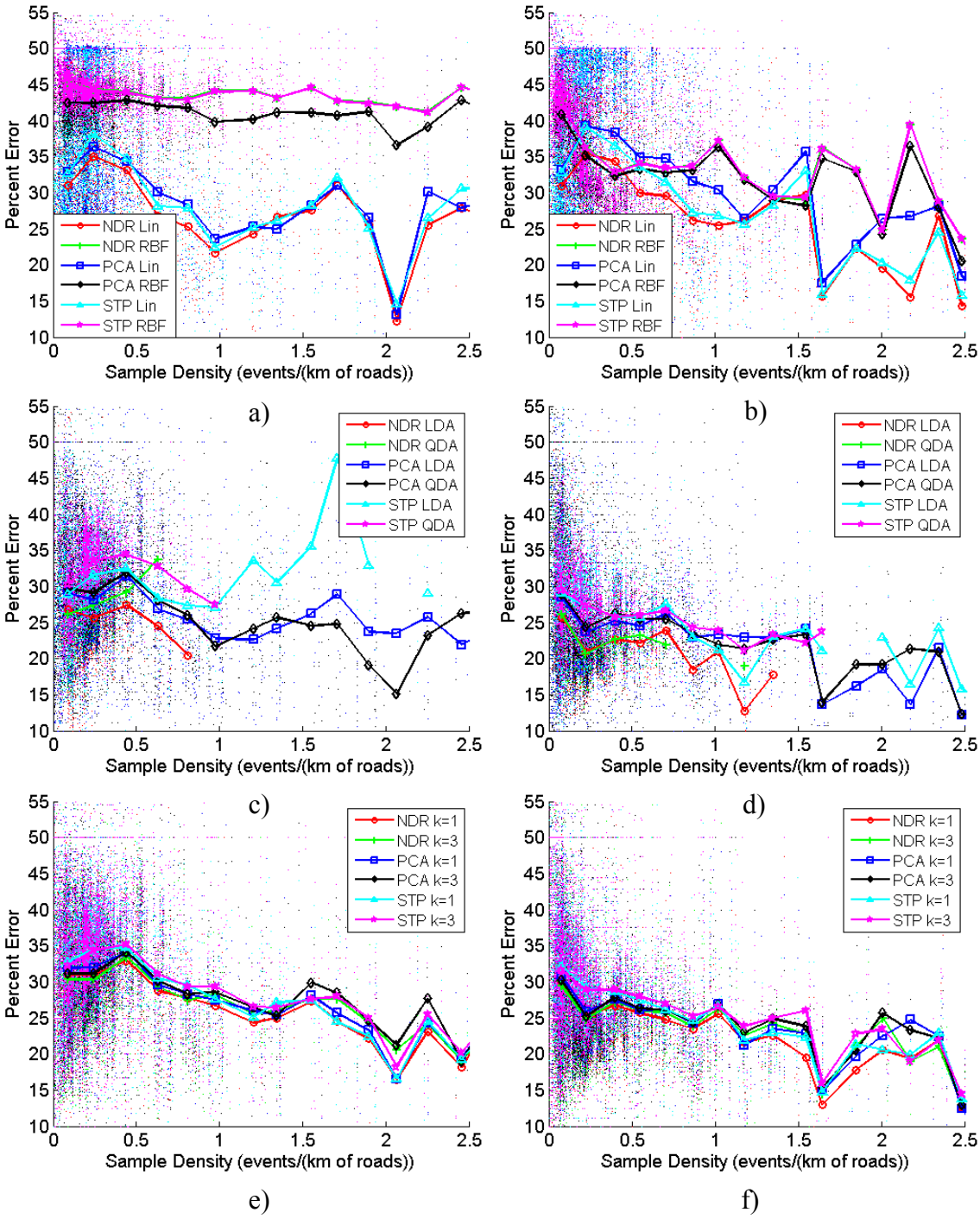
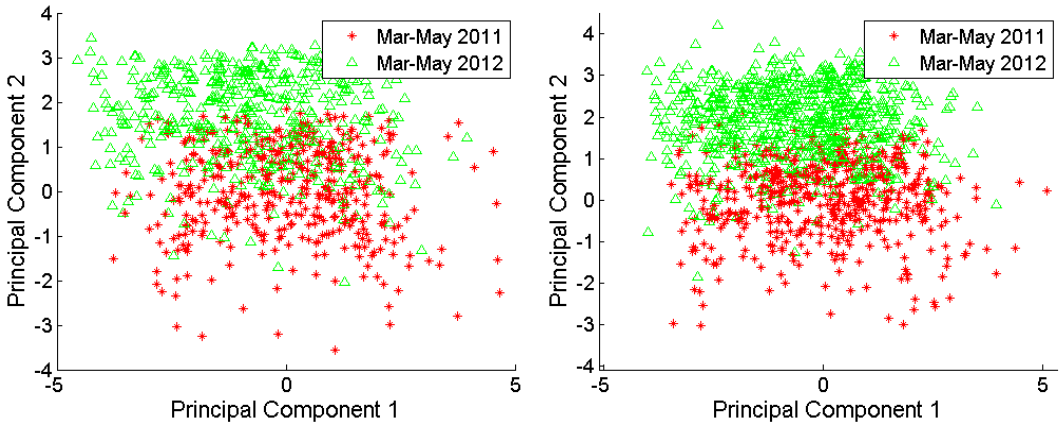



**Figure 35. The impact of sample density on classification accuracy under geographic constraints; a) IED SVM; b) DF SVM; c) IED QDA; d) DF QDA; e) IED kNN; f) DF kNN.**

directly with geography. It may also be that geographic windows alone do not sufficiently constrain the training set to produce homogeneous training instances. Across all learning algorithms, the trend displayed by SVM, fewer errors as the window radius decreases, is probably the most promising direction to investigate further.

### *Temporal Constraints*

Tactics change over time for a variety of reasons. Some changes are temporary or seasonal —tactics that require digging are difficult to accomplish when the ground is frozen in the winter— and some are permanent, for example tailored countermeasures may render some attacks obsolete. While truly temporary or transitory changes may be caused by random or unmeasured effects, seasonal patterns are relatively predictable. An example, in the case of Afghanistan, is that higher elevations tend to have fewer events in the winter. This is probably due to a combination of factors. First, the cold makes outdoor operations difficult. Any attack that requires extensive exposure to the elements is going to be difficult or even dangerous for the attacker. Also, once snow starts falling, many roads become impassible. Targets are unable to reach the higher elevations and thus, are not available to be attacked. Similar patterns also arise from religious and cultural practices. Some weekly variations may be attributed to attendance at religious events where large groups of people cluster for defined periods of time. On an annual scale, many Afghans observe the religious practice of Ramadan. During this time, some activities are prohibited or restricted. This may impact the nature and timing of some conflict events.




a) b)

**Figure 36. Comparison of conflict events over two different three-month periods in 2011 and 2012; a) IED events; b) direct fire events.**

Other more dynamic temporal patterns may exist. One example is related to availability of war materials. When the highways are passable, war materials often flow freely into conflict areas. However, if this flow is stopped or decreased, attackers may be forced to use substitutes, which constrains or changes the range of attacks available to the attacker. The cat-and-mouse game of measure/countermeasure is another example of a process that generates shifting patterns

Over time, targets increasingly thwart repeated or similar attacks through the use of countermeasures. Attackers modify their target selection and attack emplacement efforts to nullify the countermeasures. This results in a shift of attack patterns over time. Thus, adaptation to countermeasures drives changes in tactics. Whether abrupt or incremental, attackers shift and modify their patterns seeking successful attack vectors using available tactics. Successful attacks also drive pattern change. An example of this is replication of successful attacks. When an attacker is successful, the configuration

used in the attack may be attempted elsewhere. Since terrain varies, the attack configuration chosen at the new location may be similar but not identical to the previous, successful attack. If the attack is again successful, the range of suitable configurations grows and the number of possible attack sites expands.

Regardless of the reason, it is clear that tactics shift over time. Figure 36 compares tactics over two three-month periods in 2011 and 2012. In the figure, all features for IED and DF events occurring during the period March-May 2011 are normalized and then transformed using PCA. The resulting normalization factors and PCA loadings are used to transform all features of IED and DF events that occurred during the same months in 2012. For both direct fire and IED attacks, it is clear that the patterns detected in 2011 are substantially different than the patterns from 2012.

Figure 37 displays a general pattern shift over time. In the figure, all features of IED and DF events that occurred during February, 2011 are normalized and then transformed using PCA. The resulting normalization factors and PCA loadings are used to transform all features of IED and DF events that occurred at monthly intervals between March 2011 and August 2012. In the plot, the y-axis is the mean value of the first (and most significant) principal component for all events that occurred in that month. The x-axis is the date. Consecutive points of the same class are joined with a line to facilitate visual interpretation. As shown in the plot, the value of the first principal component shows significant and consistent drift over the 19 months covered in the data. Interestingly, the drift is in the same direction and roughly the same magnitude for both

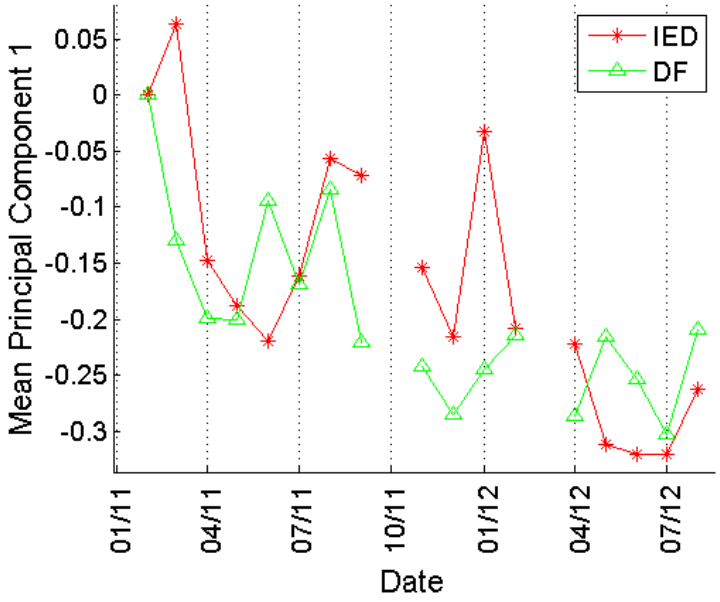



**Figure 37. IED and DF patterns shift over time.**

IED and DF events. This may indicate that both types of attacks rely on common or similar terrain features and attack site configurations.

(The gaps in October 2011 and March 2012 are caused by missing data. See Appendix A.2 for more detail.)

Figure 38 examines the impact of temporal constraints on classification accuracy for IED and DF events. For each event in *ST* and for various durations, when the total number of IED or DF events within the temporal window exceeds 10, three dimensionality reduction schemes, PCA, STP and NDR are performed on normalized data and then combined with SVM (using linear and RBF kernels), discriminant analysis (linear and quadratic) and kNN (1-NN and 3-NN) to estimate the classification error, which is calculated as the mean of the percent error. The duration of the test set is

universally constrained to 60 days. Three way cross-validation is used and error bars indicate a 95% confidence interval. (The lines connecting data points are provided to increase readability of the figure.) As shown in Figure 38, training window span or duration has an impact on classification accuracy. For DA and kNN, smaller temporal windows tend to have larger average errors with all combinations of classification scheme and dimensionality reduction generally trending together. For SVM, on the other hand, smaller windows tend to have lower classification error for linear kernels and higher error for RBF kernels. It is particularly interesting to note that smaller temporal windows have lower average error for SVM with a linear kernel. Across all combinations of classification algorithm and dimensionality reduction, the best average error rates achieved are around 30% for both IED and DF events.

As shown in Figure 38, the size of the temporal window used to produce the training set has some impact on classification accuracy. The size of the sample itself also appears to have varying when temporal windows are used. Figure 39 examines the impact of sample size on classification accuracy across all temporal window sizes. In the case of SVM with a linear kernel, small samples tend to produce the lowest average classification errors at sample sizes of around 200 events. The error rate increases with the sample size beyond this point. SVM with an RBF kernel performs poorly across all sample sizes. kNN tends to show little variation with the error rate remaining approximately flat across all sizes while DA shows an initial decrease in error up to sample sizes of 200 and generally unchanged performance after that.

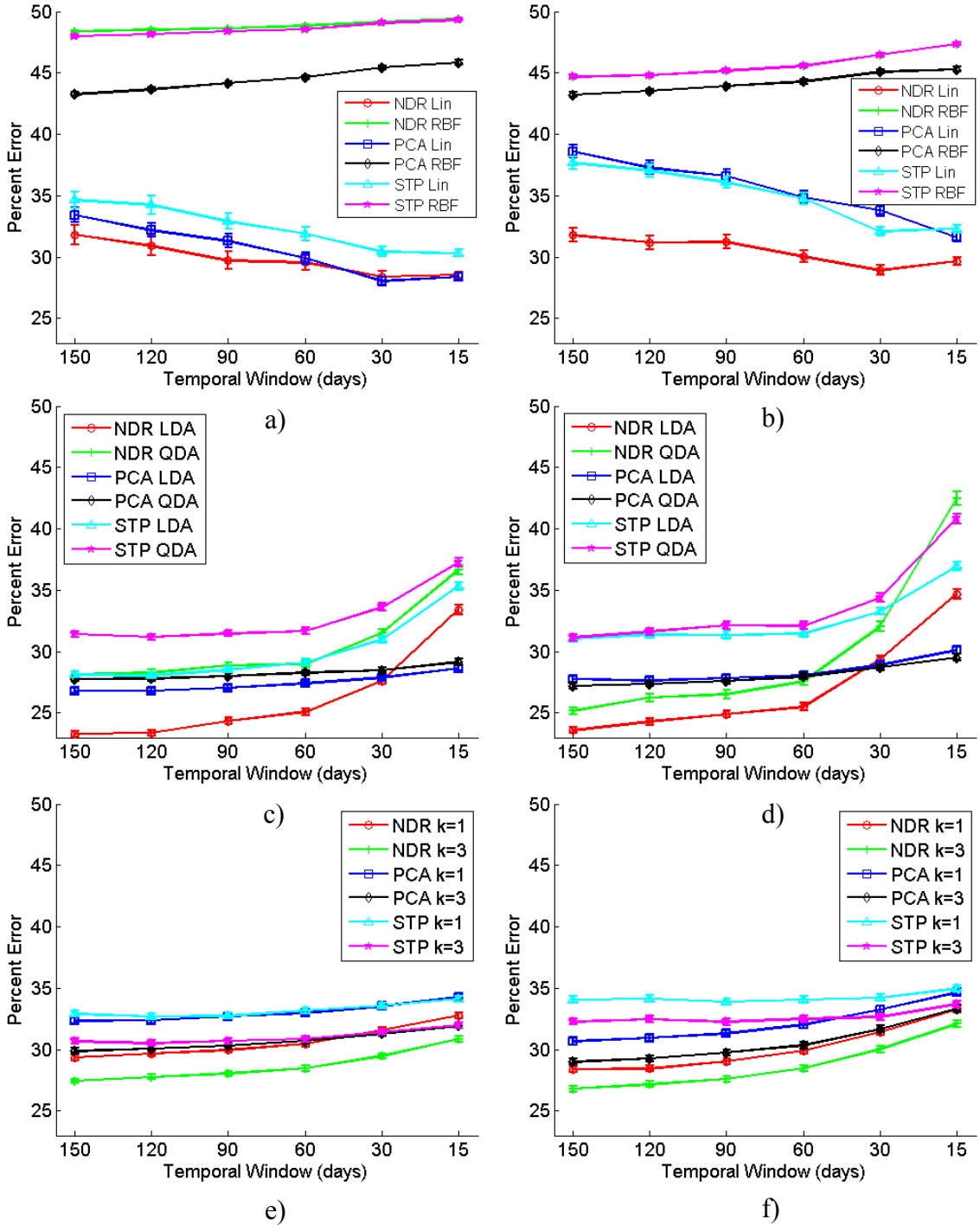



**Figure 38. The impact of the size of temporal windows on classification accuracy; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

Unlike sample size, sample density appears to have little impact on classification accuracy, as shown in Figure 40. In the figure, classification error tends to remain flat and approximately constant across all densities. The exception is DA at very low densities, but this is due to DA's previously noted problems with small samples.

Analysis of temporal constraints provides interesting insights into historical event data. The most salient insights are probably that sample size and window radius appear to have some effect on classification error. Since density appears to have little impact, the dominant effect is probably sample size. An approach to sample selection that seeks the smallest window containing some minimum sample size might improve classification accuracy. As before, SVM is the best performer at when using smaller windows, smaller sample sizes and lower density.

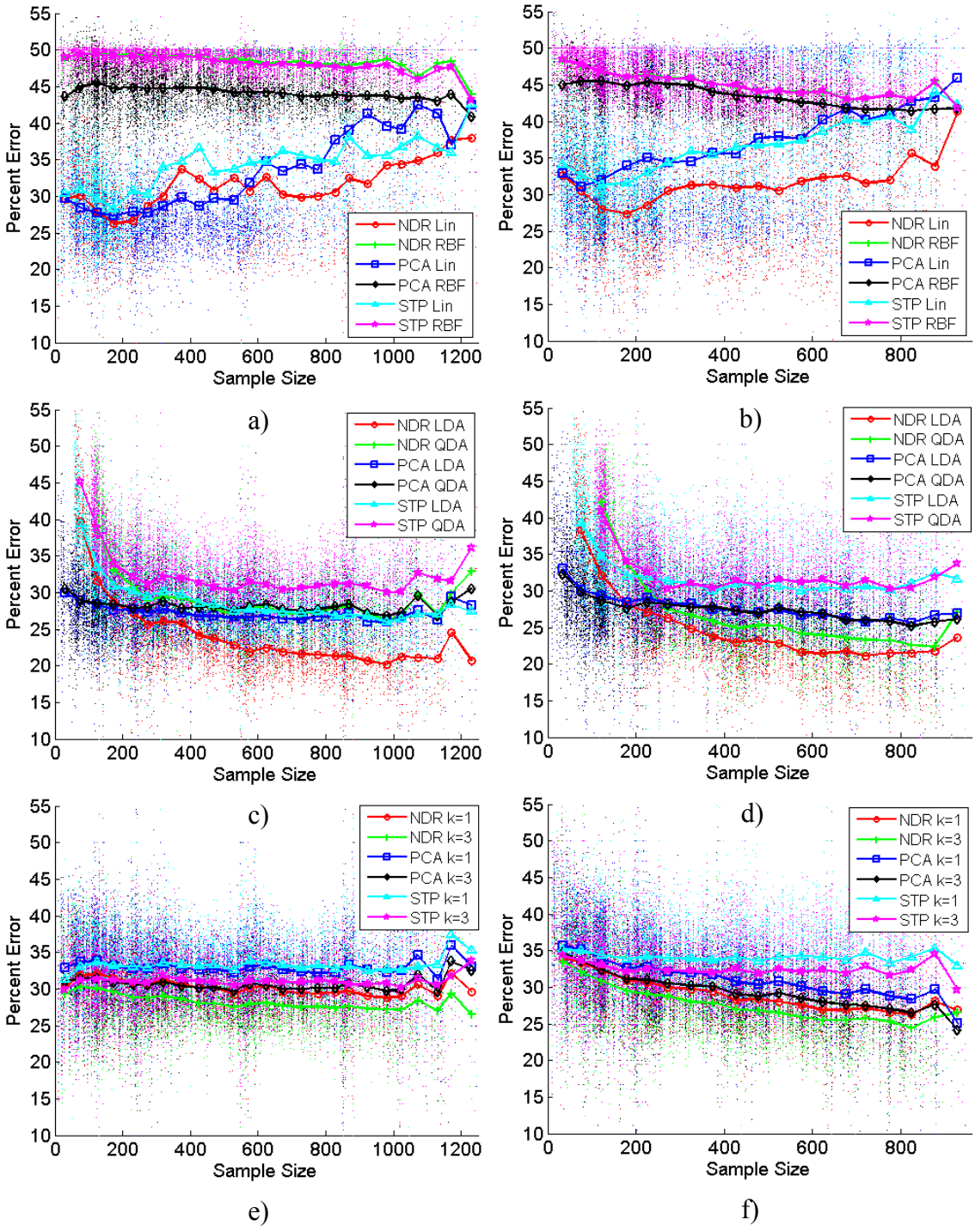



**Figure 39. The impact of sample size on classification accuracy under temporal constraints; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

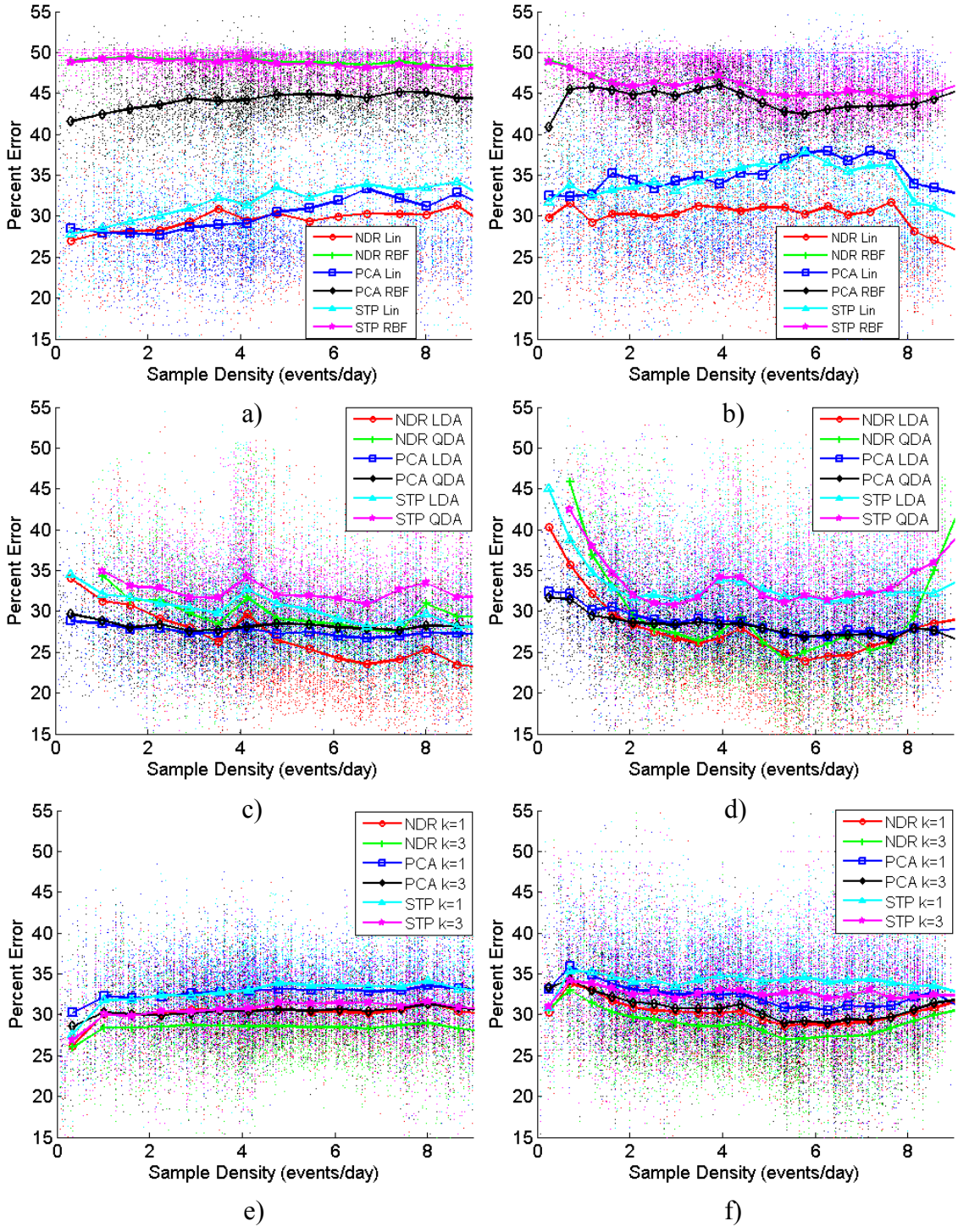



**Figure 40. The impact of event density on classification accuracy under temporal constraints; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

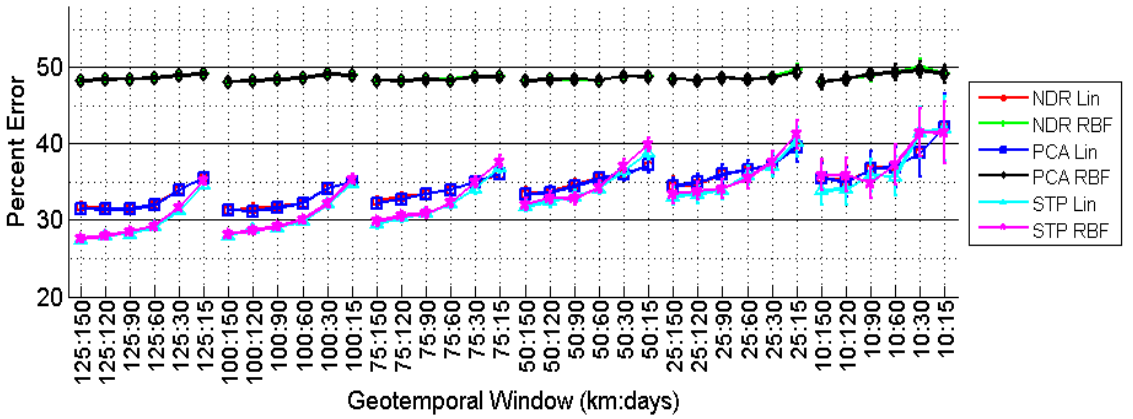
### *Combined Constraints*

Independently, both geographic and temporal constraints can be used to produce classification error rates in the vicinity of 30%. With sample sizes around 200, whether geographically or temporally selected, error rates of less than 30% are achievable. Combined geographic and temporal constraints offer a way to select training sets with the potential to produce lower error rates.

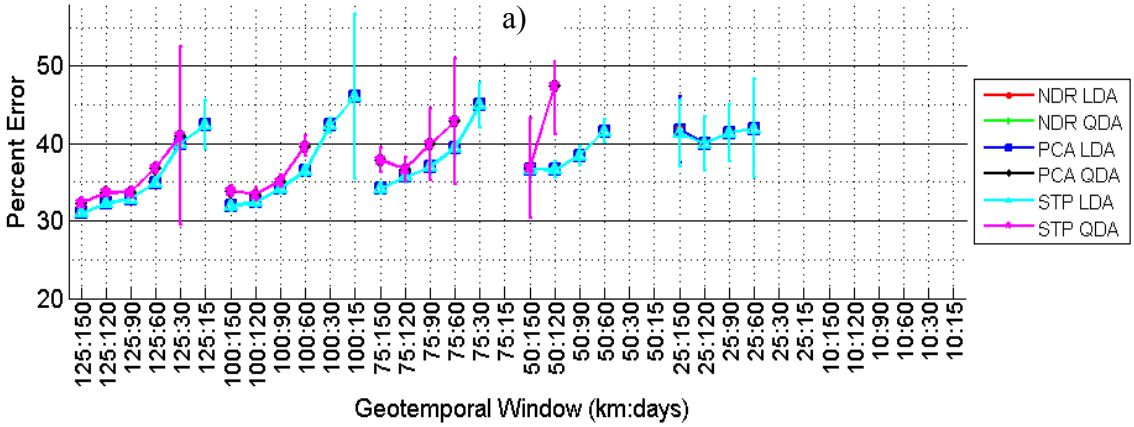
Combined constraints satisfy our intuition regarding human activities and our understanding of the measure/countermeasure cycle in modern warfare. Geographic constraints make it more likely that the attackers have something in common like group membership, shared training, access to the same war materials, or similar terrain in which to operate. Temporal constraints make it more likely that the same tactics are in use, the same or similar war materials are available to attackers, and the same countermeasures are shared by targets. When the two types of constraints are combined, the resulting datasets are probably likely to be more homogeneous and probably more likely to share a common and more regionally focused set of tactics.

Figure 41 examines the impact of combined geographic and temporal constraints on classification accuracy for IED events. (The plots for DF events are similar and can be found in Appendix D.)

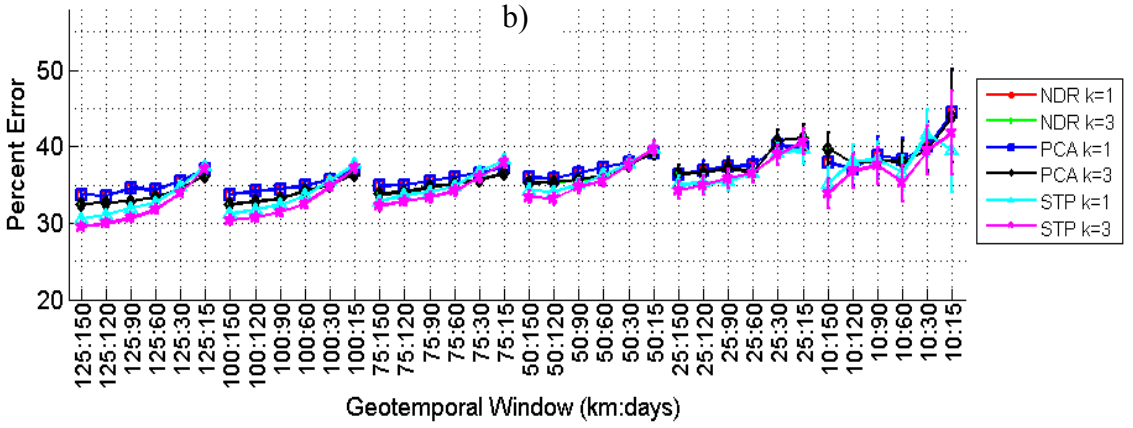
For each event in *ST* and for various durations, when the total number of IED events within the combined geographic and temporal training windows *b* exceeds 10,




a)




b)




c)

**Figure 41. The impact of combined geographic and temporal windows on classification accuracy of IED events; a) IED SVM; b) IED DA; c) IED kNN.**

three dimensionality reduction schemes, PCA, STP and NDR, are combined with SVM (using linear and RBF kernels), discriminant analysis (linear and quadratic) and kNN (1-NN and 3-NN) to estimate the classification error, which is calculated as the mean of the percent error. The duration of the test set is universally constrained to 60 days. Three-way cross-validation is used and error bars indicate a 95% confidence interval. In the figure, the x-axis is the < geographic window : temporal window > combination used to selecting the training set for the machine learning algorithm. Note that the plots are grouped by geographic constraint, i.e. 100 km, with a break in the connecting line signifying the jump to the next geographic group. (The lines connecting data points are provided to increase readability of the figure.)

In Figure 41, classification accuracy is between 30-40% for both SVM and kNN with SVM performing slightly better. As before, SVM with linear kernels perform significantly better than SVM with RBF kernels. Note that these are still using the default parameters for the SVM box constraint and RBF scaling factor. Both SVM and kNN show fairly constant accuracy within each geographic constraint group with an upwards trend in error of approximately 3-5% as the temporal window shrinks. The increase in classification error is more significant for DA and it continues to have problems with small datasets. This is reflected in the missing error values for classification attempts using NDR and PCA.

Figure 42 shows the impact of sample size on classification accuracy under geographic and temporal constraints. In general, classification accuracy improves as the sample size increases. SVM shows the best performance at sample size around 200

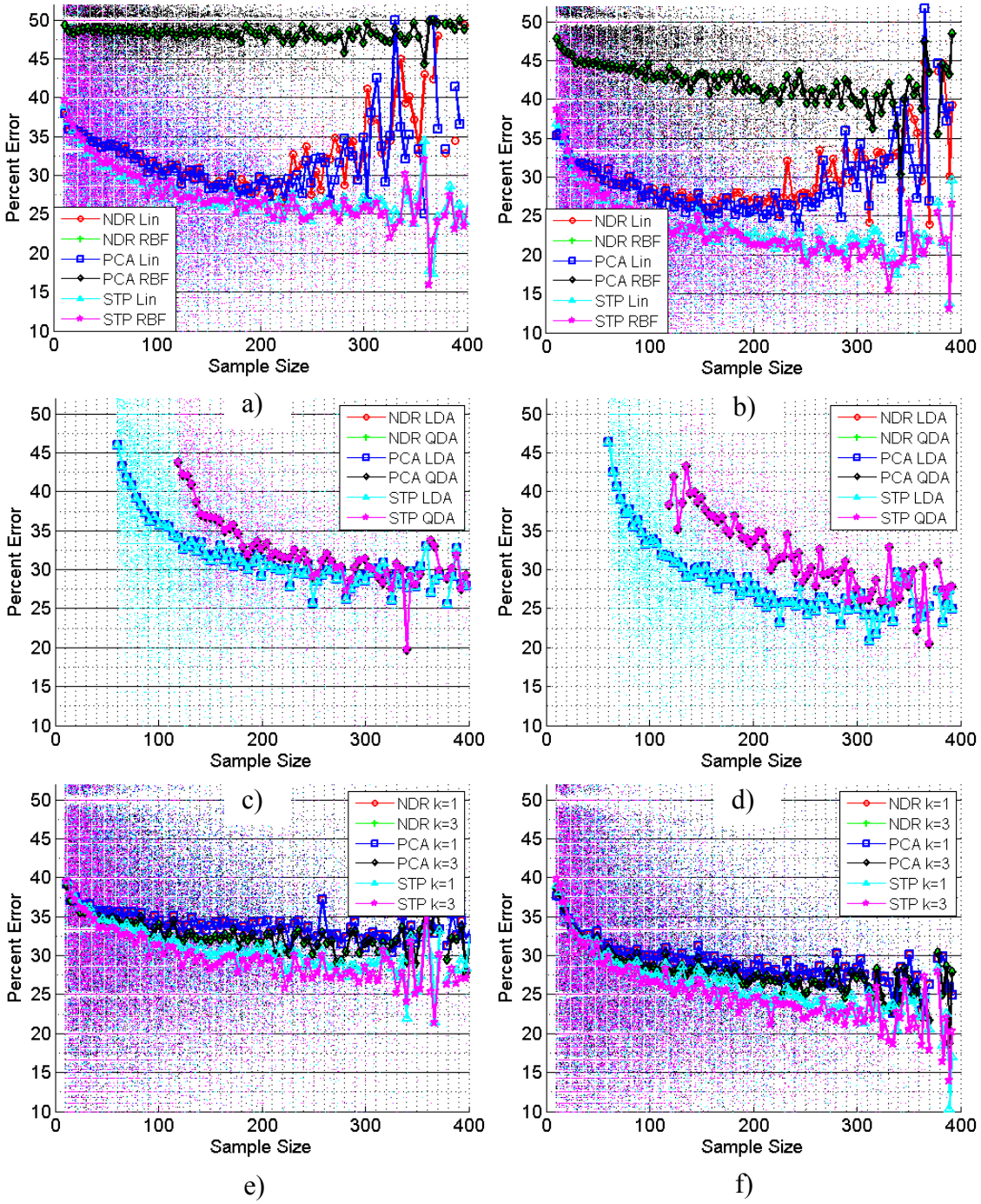



**Figure 42. The impact of sample size on classification accuracy under combined geographic and temporal constraints; a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

while kNN has its best performance at the largest sizes. DA's classification accuracy is similar to that of SVM and kNN for large samples. Small samples continue to be problematic for DA.

### Feature Selection and Dimensionality Reduction

In the previous experiments involving geographic, temporal and geotemporal constraints, default settings were used for PCA and stepwise feature selection. In the case of PCA, all principal components were used, regardless of their contribution. Similarly, for stepwise feature selection, all weighted features were used. However, it may be possible to reduce error by reducing dimensionality. With PCA, one method of reducing dimensionality is by assessing the amount of variance accounted for in the reduced model. The number of principal components in the final model is controlled by limiting the total cumulative variance explained by these components. For stepwise feature selection, dimensionality may be managed by varying the *p*-value threshold (*penter* parameter). Smaller *p*-values lead to smaller models

Figure 43 examines the impact of varying cumulative variance (for PCA) and the *penter* parameter (for STP). For conciseness, only three combinations of learning algorithm and dimensionality reduction are applied to the IED data and shown here. Both PCA and STP are combined with SVM using a linear kernel, LDA, and kNN with k=1. For each combination, the error rate produced by cumulative variances between

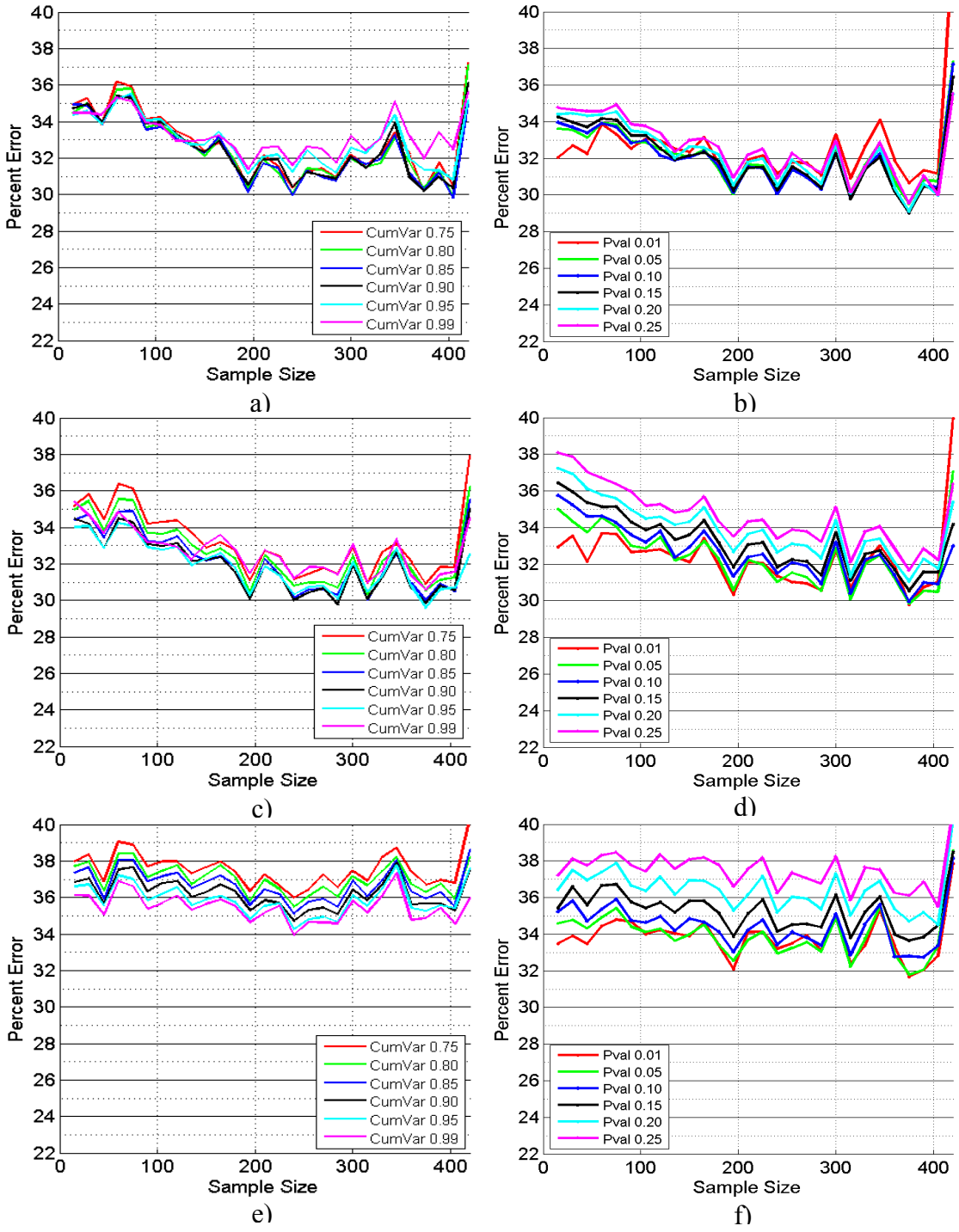



**Figure 43. The impact of varying cumulative variance and *penter* parameters on IED classification; a) PCA with SVM linear kernel; b) STP with SVM linear kernel; c) PCA with LDA; d) STP with LDA; e) PCA with kNN (k=1); f) STP with kNN (k=1).**

0.75 (75%) and 0.99 (99%) and *penter* parameters between 0.25 and 0.01 are shown. A complete set of plots for both IED and DF data is available in Appendix D.

For PCA, changing the cumulative variance appears to have little effect using SVM, LDA or kNN. For all three of these learners, a cumulative variance of 95% showed consistently good performance across the entire range of sample sizes. For kNN, a cumulative variance of 0.99 performed best but this performance was not shared by SVM and LDA.

For STP, a *penter* parameter of 0.01 consistently produces the lowest error rates at small sample sizes but performs less well at large sample sizes. However, a *penter* value of 0.05 performs well across the entire range of samples sizes. Error rate differences as high as 5% were seen, with larger *penter* values tending to produce higher error rates, especially at smaller sample sizes.

Figure 44 displays the number of features or principal components in the final model after dimensionality reduction is performed. In the case of PCA, the largest model results from including or accounting for the largest amount of variance in the data. A significant increase in model size occurs between 95% and 99% cumulative variance. For STP, the largest models are produced with the largest or most relaxed *penter* value and the size of the final model changes roughly uniformly as *penter* is varied. One interesting note is that the sample size appears to have little impact on the size of the final model. For PCA, smaller samples tend to produce slightly smaller models. For STP, there seems to be little effect.

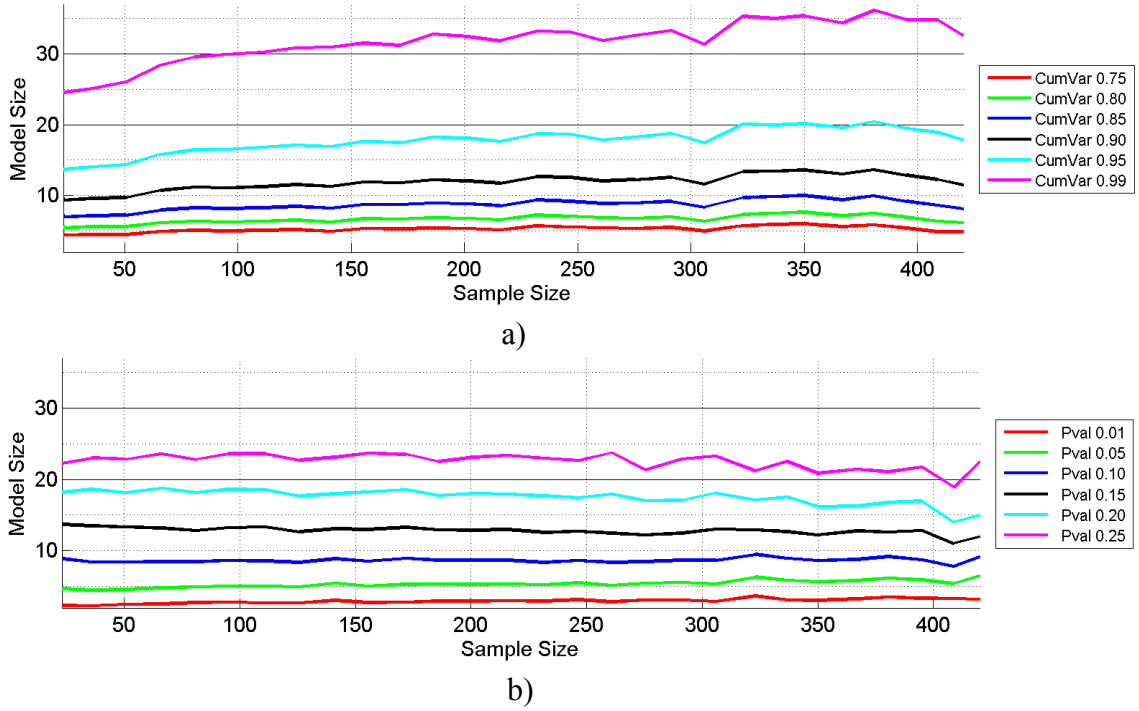



**Figure 44. Model size after dimensionality reduction; a) PCA; b) STP.**

Based on the results shown in figures 43 and 44, a cumulative variance of 95% will be used for the experiments in this research that involve PCA. A *penter* value of 0.05 will be used for experiments that involve STP. Interestingly, these values are the defaults in Matlab's *stepwisefit* function.

For a given tactic or type of attack, some features are more relevant and tend to be selected more often by stepwise feature selection. Figure 45 shows the selection rate of features used in the experiment that generated the data in Figure 43. Of the 77 features, only 13 are selected more than ten percent of the time. An additional 20 features are selected between 5-10% of the time. The two most frequently selected features are elevation (1) and proximity to a human population center of any size (58).

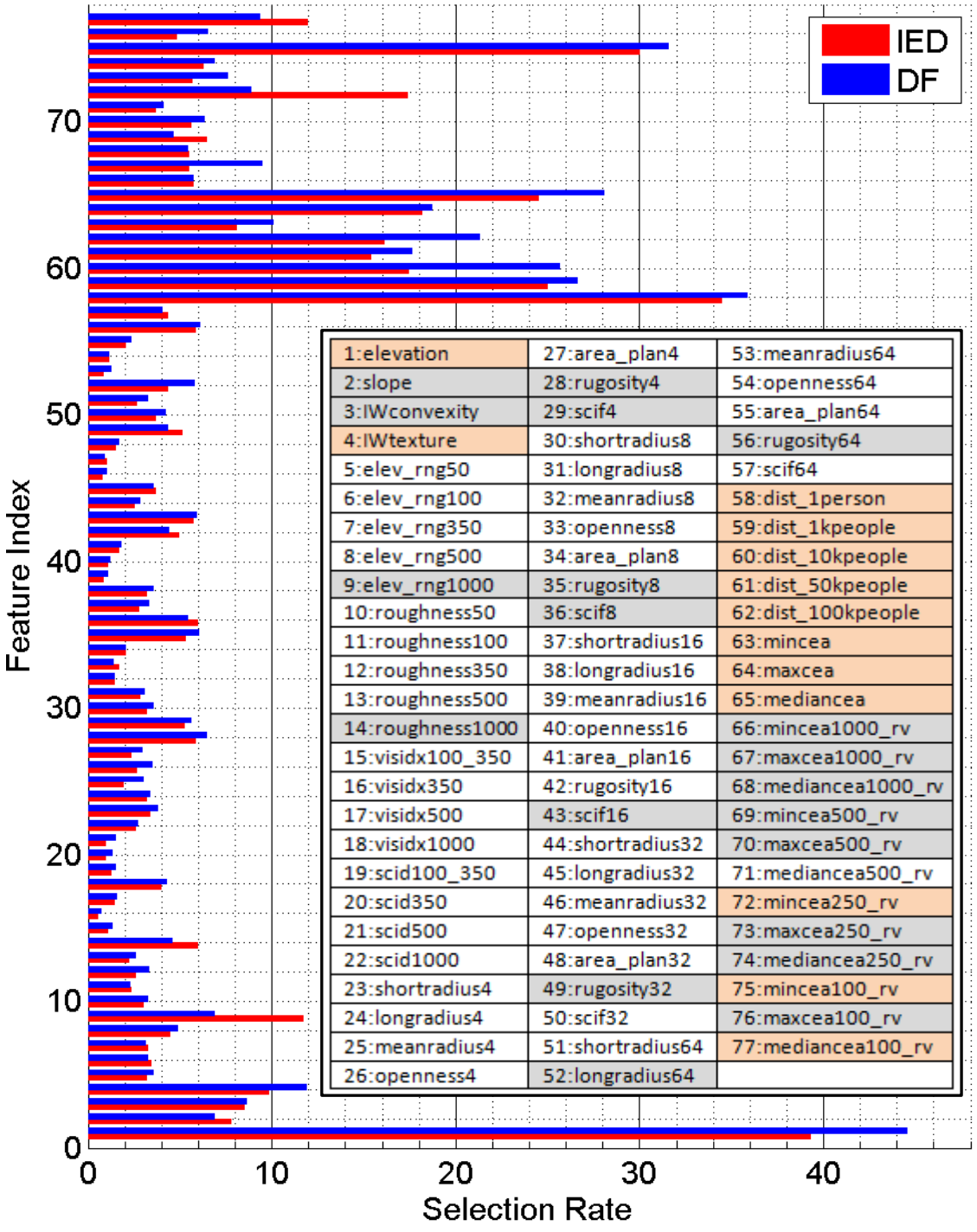



**Figure 45. Features selected using stepwise selection for dimensionality reduction. The inset lists the features by name and highlights the largest contributors: light gray boxes are selected in 5-10% of all classification attempts, light peach boxes are selected in > 10% of all classification attempts.**

Two observations emerge. First, the features contributing to the classification of IED and DF events are almost always selected at roughly the same rate. This may mean that both types of events can be treated as a single class Also, the 33 features most frequently selected by STP fall into a few categories. Fourteen of the features are related to cumulative escape adjacency (CEA) (denoted as 'cea' in the Figure 45 inset) and a visibility metric related to CEA. Five of the features are related to the distance from human population centers. Radial-based shape complexity and radial-based rugosity both appear at four different resolutions.

Figure 46 examines the impact on classification accuracy for IED events when using PCA and stepwise feature selection to reduce the dimensionality of datasets selected using combined geographic and temporal constraints. (The plots for DF events are similar and can be found in Appendix D.) For each event in *ST* and for various durations, when the total number of IED events within the combined geographic and temporal windows exceeds 10, normalized data is used with three dimensionality reduction schemes, PCA, STP and NDR, and combined with SVM (using linear and RBF kernels), discriminant analysis (linear, quadratic) and kNN (1-NN, 3-NN) to estimate the classification error, which is calculated as the mean of the percent error. The duration of the test set is constrained to 60 days. Three-way cross-validation is used and error bars indicate a 95% confidence interval. In the figure, the x-axis is the <geographic window : temporal window> used for training the machine learning algorithm. Note that the plots are grouped by geographic constraint, i.e. 100 km, with a break in the

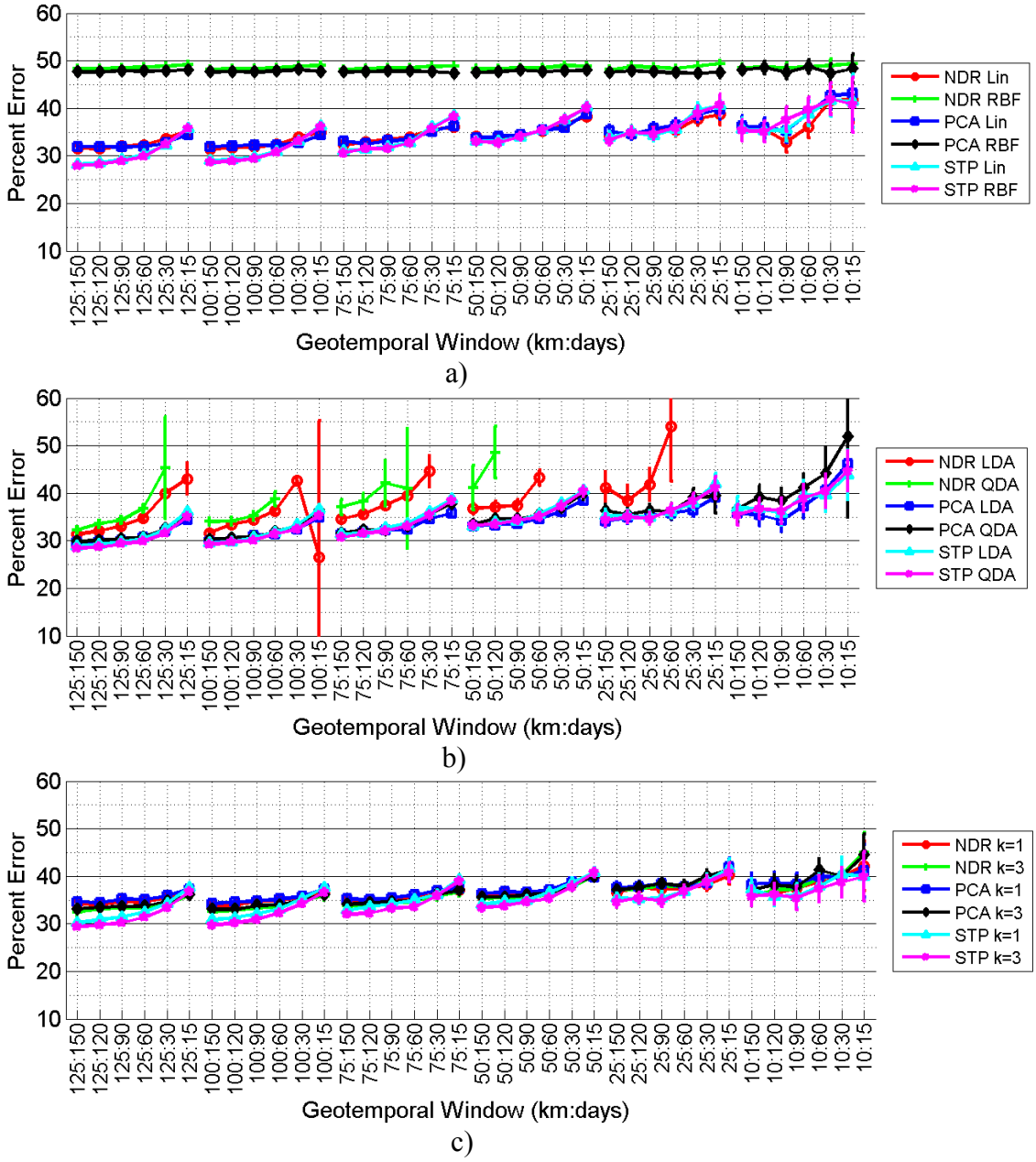



**Figure 46. The impact of dimensionality reduction using PCA (cumulative variance <= 95%) and STP (*penter* <= 0.05) on classification accuracy of IED events when using combined geographic and temporal windows; a) IED SVM; b) IED DA; c) IED kNN.**

connecting line signifying the jump to the next geographic group. (The lines connecting data points are provided to increase readability of the figure.)

In Figure 46, classification accuracy is around 30% for both SVM and kNN. As before, SVM with linear kernels performs significantly better than SVM with RBF kernels. Note that these are still using the default parameters for the SVM box constraint and RBF scaling factor. Both SVM and kNN show fairly constant accuracy within each geographic constraint group with an upwards trend in error of approximately 3-5% as the temporal window shrinks. The increase in classification error is similar for DA. This is because dimensionality reduction contributes to overall improved performance for DA. With reduced dimensionality, more of the datasets have full rank and can be successfully inverted. DA still fails on most small data sets.

Figure 47 shows the impact of sample size on classification accuracy under geographic and temporal constraints when the cumulative variance of the PCA components is constrained to be $\leq 95\%$ of total variance and the maximum p-value of the stepwise-selected features is constrained to 0.05. Across all classification methods, larger sample sizes tend to produce lower error rates with the lowest rates around 25%. DF tends to have consistently lower error rates that IED, although the difference is not great. It is interesting to note that stepwise feature selection consistently outperforms PCA across all classification methods.

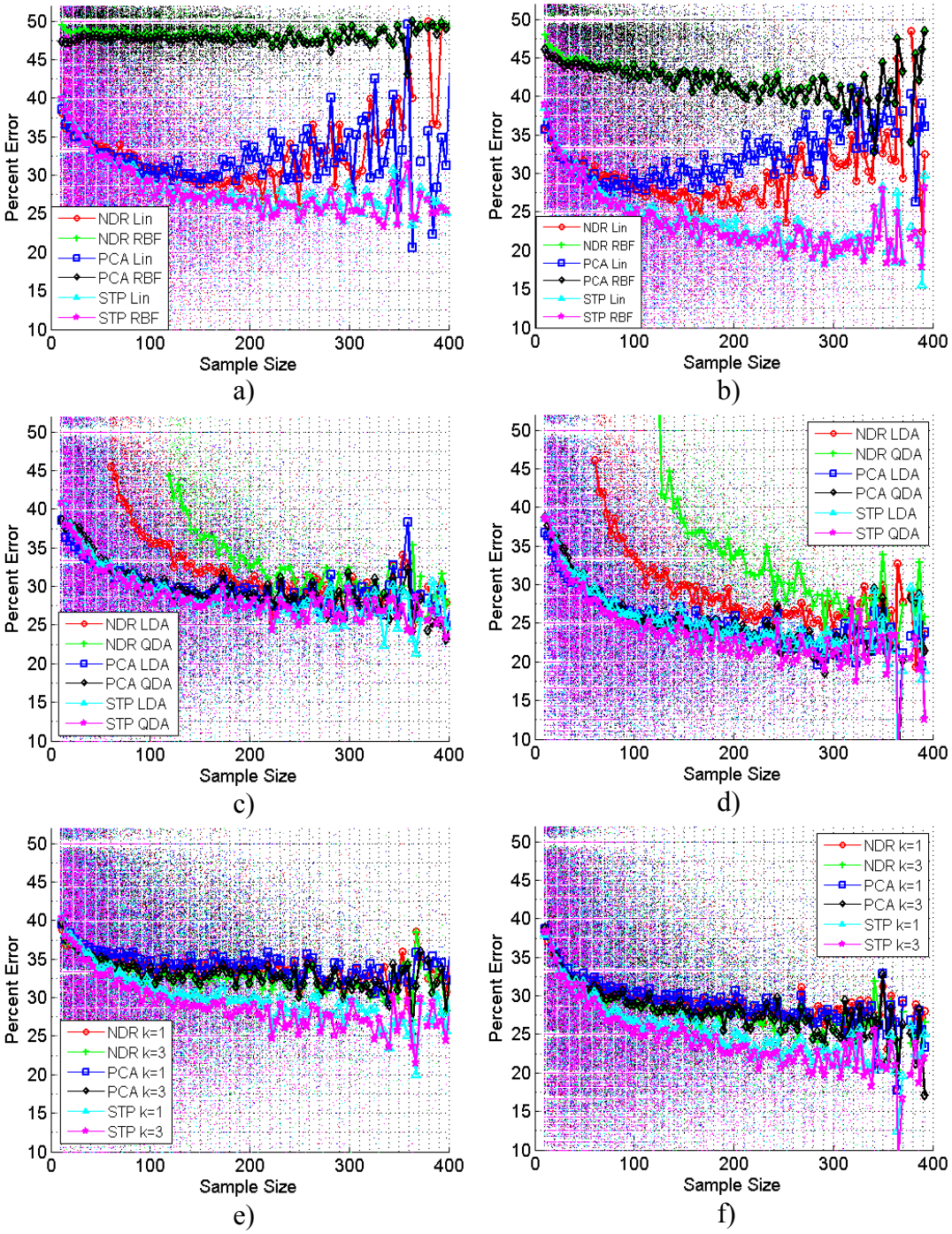



**Figure 47. The impact of sample size on classification accuracy under combined geographic and temporal constraints when the cumulative variance of the PCA components ≤ 95% and the maximum *p*-value of the stepwise-selected features ≤ 0.05. a) IED SVM; b) DF SVM; c) IED DA; d) DF DA; e) IED kNN; f) DF kNN.**

## Parameter and Hyperparameter Estimation

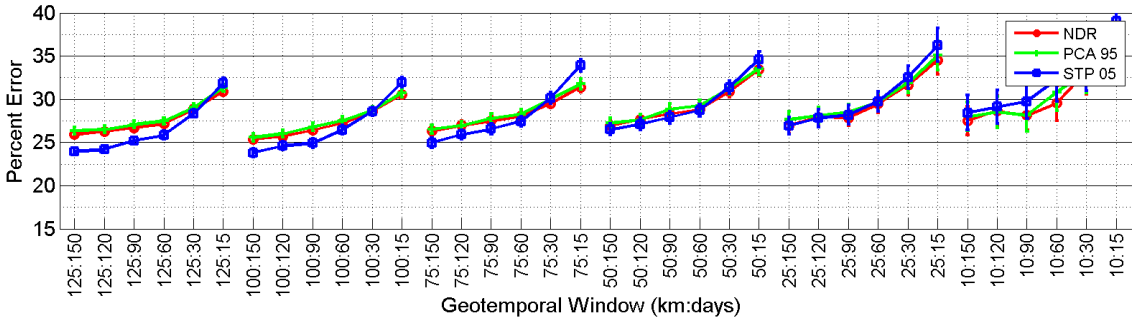
The previous section showed that geotemporal constraints and dimensionality reduction combine to produce error rates on the order of 25% with large sample sizes. Stepwise feature selection produced the lowest average error rates in Figure 46 across all combinations of dimensionality reduction and classification method examined here. However, the parameters and hyper-parameters used were set to default values. In particular, both the SVM box constraint and the SVM RBF kernel shape parameter were set to value of 1. For kNN, the number of nearest neighbors was fixed to values of 1 and 3. Dynamic estimation of these parameters may improve classification accuracy.

### *Dynamic Estimation of k for kNN*

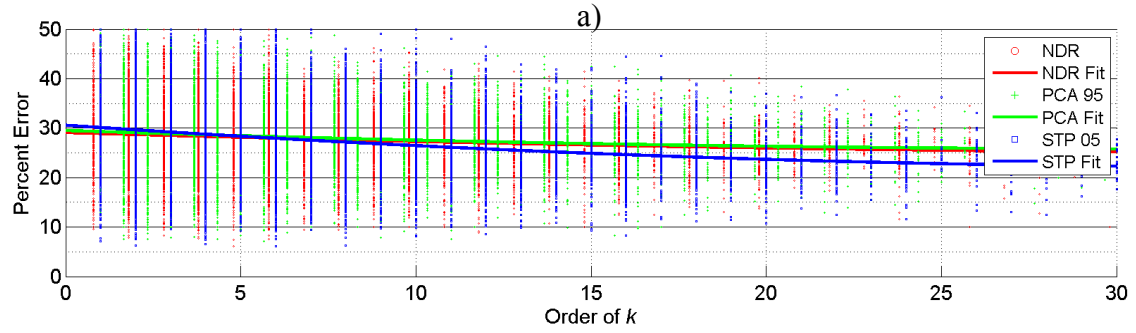
The order of $k$ in kNN has a strong influence on classification error. In [81], the authors cite two examples that demonstrate this effect. In the first example, setting $k$=25 produced a misclassification rate of approximately 15% while $k$=1 produced a misclassification rate of approximately 25%. On a different data set of the same size and same approximate distribution of classes, the opposite was found with the lowest misclassification rate at $k$=1. In both cases, proper selection of $k$ reduced the misclassification rate significantly. In [144], Enas and Choi note a number a factors influencing the choice of $k$. For both balanced and unbalanced datasets ranging in size from 50 to 354 samples, they note that optimal vales of $k$ tend to decrease as the population correlation structures become more dissimilar. Optimal values of $k$ fluctuate with sample size and show a generally improved classification performance as the size of the dataset grows.

A variety of methods have been employed to empirically estimate $k$ from available data. Enas and Choi explore an estimation scheme optimized for small to moderate samples sizes that empirically selects the best performing $k$ selected from the range $k = n^{i/8},\ \ i = 1, \ldots, 5$, where $n$ is the size of the dataset. They find that the best results are found at $i$ = [2, 3]. Hall, et al. [145] empirically select an optimal $k$ by finding the some $k$ that produces the lowest error rate and then adjusting for sample size and dataset dimensionality. In [146], the value of $k$ is affected by the size of the smallest class in the dataset. Ghosh [143] proposes a spatially adaptive selection scheme where $k$ lies in the range $\left[1\ C\sqrt{n}\right]$. The size of the sample is $n$ and $C$ is a constant, set to 2 in Ghosh's research. $C\sqrt{n}$ is constrained to be no larger than the size of the smallest class. Common to all of these methods is an iterative step where the classification accuracies for different values of $k$ are compared. The smallest misclassification rate is used to determine $k$.

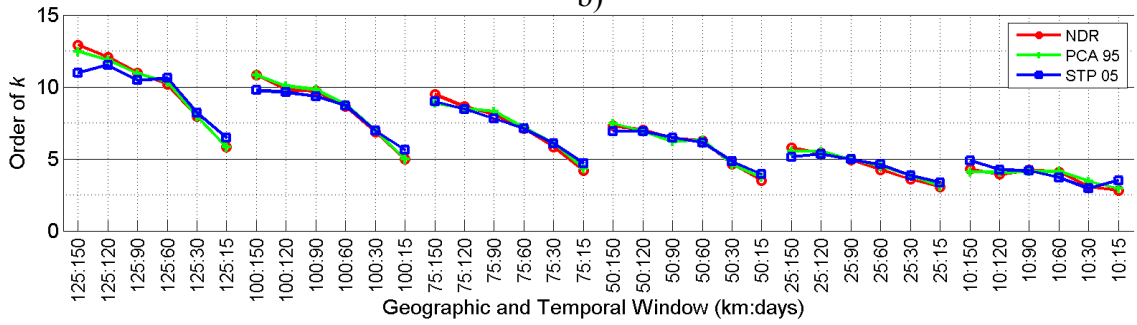
In this research, we bound the upper value of $k$ to be $3\sqrt{n}$, and further constrain this upper bound to be no larger than the size of the smallest class. Its optimal value is determined by exhaustively testing all odd $k$ in the range $\left[1\ 3\sqrt{n}\right]$. Figure 48 examines the impact of dynamically selecting $k$ on the classification of IED events. A similar figure for DF events is available in Appendix D. In Figure 48.a, the dynamic selection of $k$ decreases the misclassification rate to approximately 25%. The misclassification rate climbs as the window gets smaller, likely a reflection of smaller sample size. Stepwise feature selection slightly outperforms the other dimensionality reduction schemes. Figure 48.b examines the impact of the order of $k$ on classification accuracy. The solid



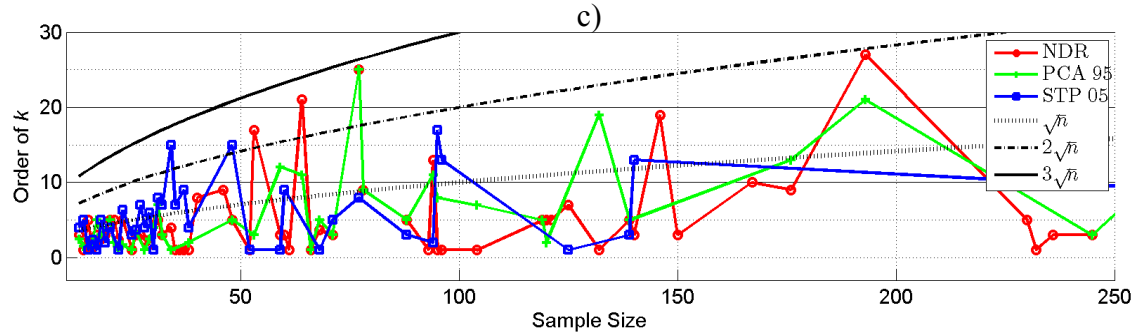

a)




b)




c)




d)

**Figure 48. The impact of dynamically varying the order of *k* on the classification of IED events using kNN; a) Classification accuracy under combined geographic and temporal constraints using varying *k*; b) Classification accuracy at various values of *k; c)* Mean order of *k* at various window sizes; d) Mean order of *k* at various sample sizes.**

lines are a second-order polynomial fit to the available data for each classification method. As *k* increases, the classification rate improves slightly. Once again, this is likely to be a reflection of sample size. Figure 48.c shows how the order of *k* changes with window size. In the figure, the median value of *k* is presented for each geotemporal window. Interestingly, optimal values of *k* tend to be small. Figure 48.d examines the impact of sample size on the order of *k*. Each point is the average of all classification attempts at that sample size, regardless of window. The results are line with the results reported by Ghosh regarding the upper bound and choice of C, with most *k* at or below $2\sqrt{n}$ and all *k* below $3\sqrt{n}$.

*Dynamic Estimation of SVM Hyperparameters*

Up to this point, SVM has been implemented using one or two fixed hyperparameters. The box constraint, also described as the cost parameter [81], penalty factor [147][104] and regularization parameter [97], was set to a default, fixed value of 1 for SVM using both linear and RBF kernels. For SVM using an RBF kernel, the shape parameter for the radial basis function was also set to a default, fixed value of 1. However, these fixed values are not appropriate for every data set and optimal values may vary with the dataset or even the specific learning.

**Box Constraint Selection**

To select the box constraint for SVM using a linear kernel, we implement an exponential search in the range $C = 2^i$ for $i=[-5\ 5]$, in 0.25 increments. This approach is implemented as a two-step process employing a coarse grid using increments of 1 and a fine grid with increments of 0.25 for optimization. The selected *C* produces the lowest

classification error using SVM with a linear kernel and three-way cross-validation.

Figure 49 examines the impact of the size of the box constraint *C* on classification accuracy when using SVM with a linear kernel. By selecting a more optimal *C* than the default of 1, classification error approaching 23% can be achieved for IED events, as noted in Figure 49.a. Figure 49.b shows that the order of *C* has little impact on classification accuracy. Figures 49.c shows that *C* varies little for NDR and PCA with respect to window size. For STP, the trend is to decrease as the window shrinks. A similar outcome can be seen in figure 49.d, where the order of *C* is fairly constant for NDR and PCA. The order of *C* using STP tends to increases with increasing sample size. Note that Figure 49.c has no log(C) values greater than 0 while Figure 49.d has log(C) values greater than zero. This is an impact of sample size. Larger windows tend to produce larger sample, but not always. Some of the samples are smaller and tend to produce smaller values of C. The results for DF events can be found in Appendix D.

**Radial Basis Function Box Constraint and Shape Parameter Selection**

A variety of methods have been proposed to select the RBF shape parameter, σ, including a grid search [97], Quasi-Newton optimization [147], and transformations of interclass distances [104]. In this research, we examine the classification accuracy obtained by selecting σ using two methods: grid search across a set of fixed values and direct estimation based on class separability in the kernel space as proposed by [106].

The grid search method uses exponential growing sequences for the box constraint, *C*, and linear search between fixed bounds for σ, as demonstrated in [98]. The range of σ is constrained to $\sigma=[0.25\ 30]$, bounds that were empirically determined from

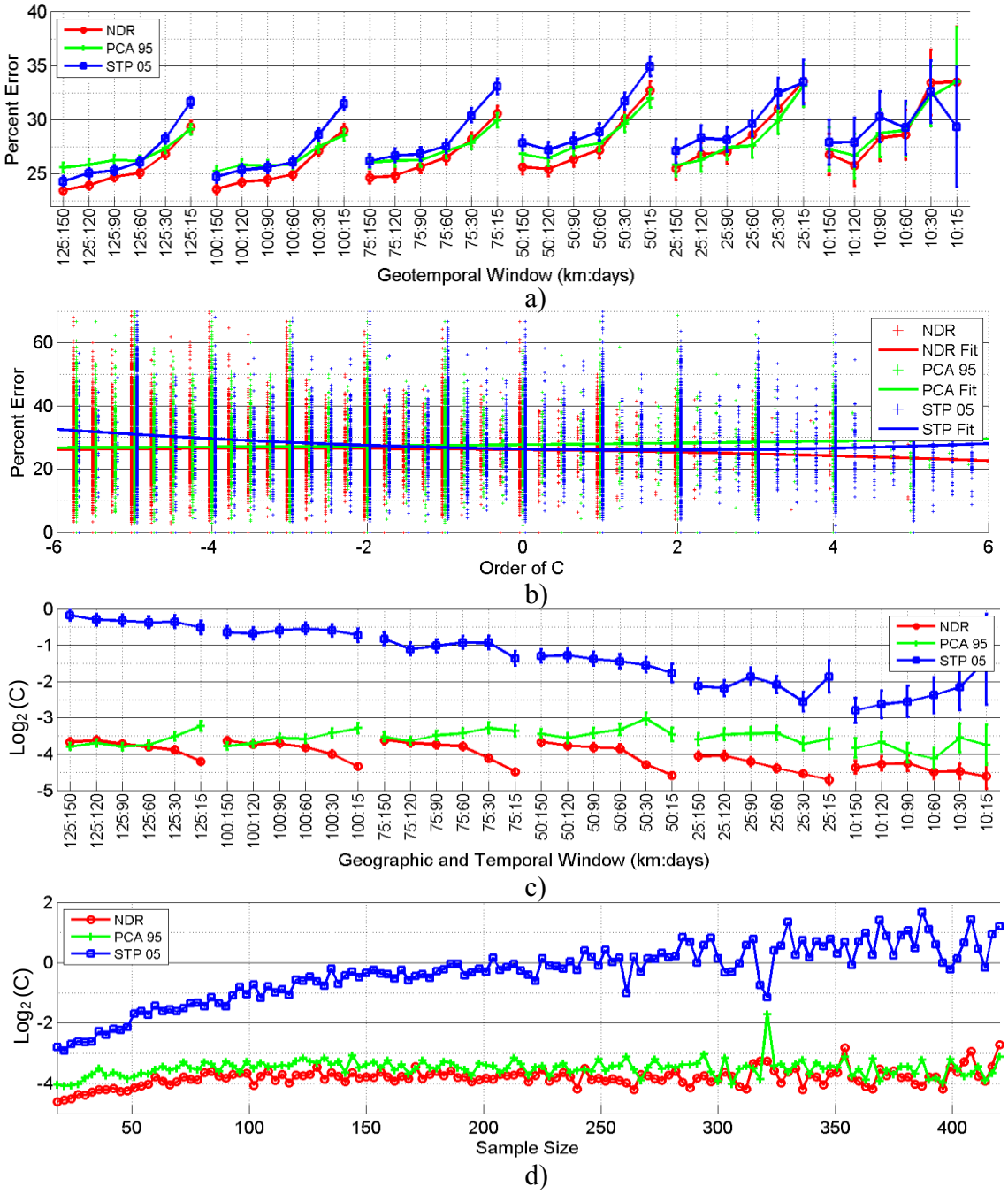



**Figure 49. The impact of dynamically varying the box constraint *C* on the classification of IED events using SVM with a linear kernel; a) Classification accuracy under combined geographic and temporal constraints using varying *C*; b) Classification accuracy at various values of *C*; c) Mean order of *C* at various window sizes; d) Mean order of *C* at various sample sizes.**

analysis of the data. The grid search was implemented as a two-step process using a coarse grid first and then refining the results using a fine grid. The outcome of the grid search for a given set of samples is the *[C,σ]* pair that produces the lowest classification error.

It is also possible to estimate σ based on class separability in the kernel space. Proposed by Liu and Zuo in [106], the value of σ is a function of between-class and within-class separability, tuned to maximize between-class separability. The authors note that in Gaussian RBF kernel space, the norm of any instance is equal to one:

$$\|\Phi(\boldsymbol{x}_i)\|^2 = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_i) = 1 \tag{61}$$

, where $\kappa(\boldsymbol{x}_i, \boldsymbol{x'}_i) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x'}_i\|_2^2}{2\sigma^2}\right)$

Thus, distance similarity and cosine similarity can be expressed as

$$\left\|\Phi(\boldsymbol{x}_i) - \Phi(\boldsymbol{x}_j)\right\|^2 = 2 - 2 \cdot exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\sigma^2}\right) \tag{62}$$

$$cos\theta\left(\Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j)\right) = exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\sigma^2}\right) \tag{63}$$

Based on the property described in Equation 61, the similarity measures noted in Equations 62 and 63, the within-class mean distance *W'*, and the between-class mean distance *B'*, Liu and Zuo propose an estimate of σ defined as

$$\hat{\sigma} = \sqrt{\frac{B' - W'}{4 \cdot \log(B'/W')}} \tag{64}$$

Figure 50 notes the impact of varying *C* and *σ* on classification of IED events across a variety of geotemporal windows. In the figure, the cumulative variance of the

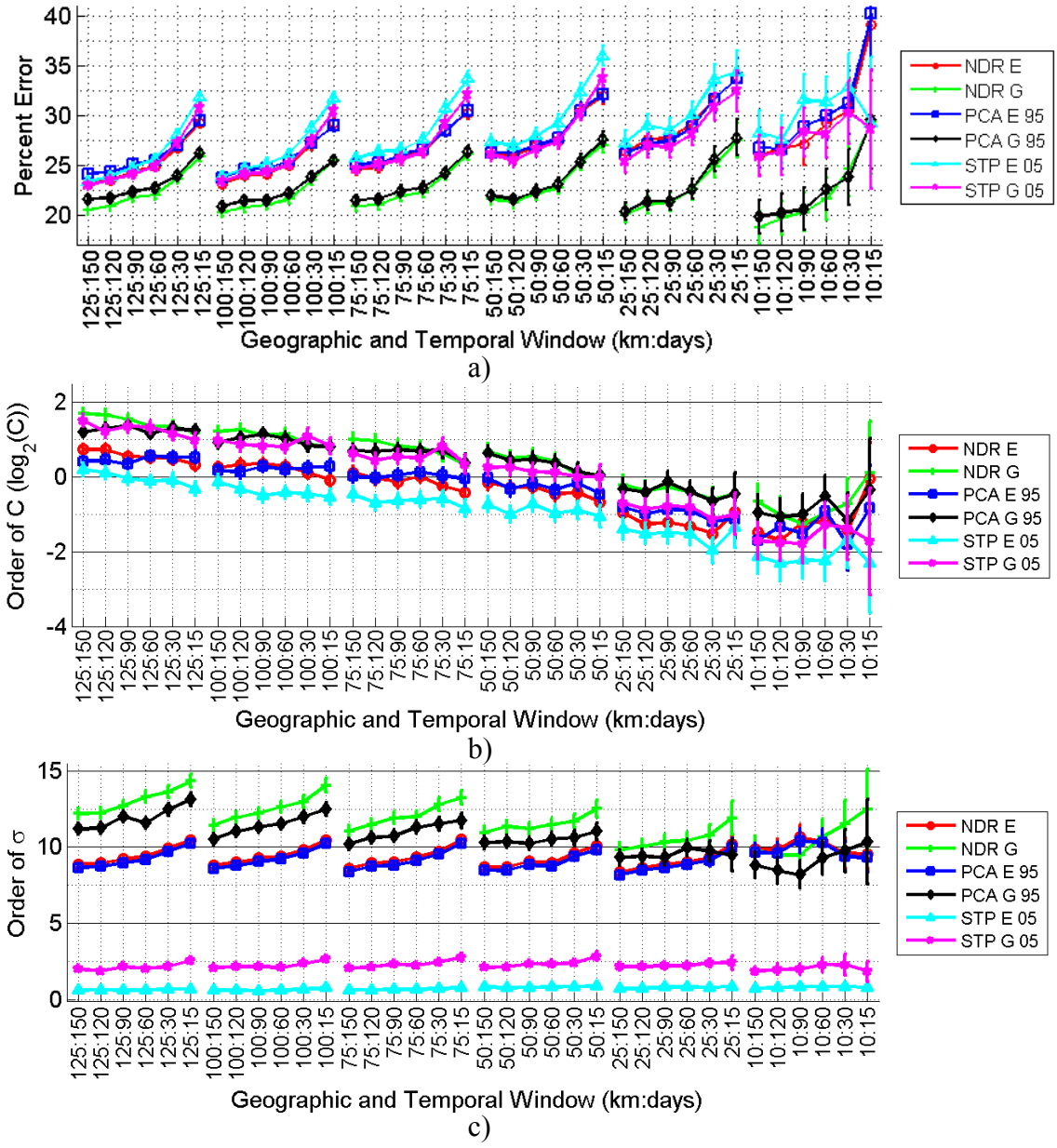



**Figure 50. The impact of dynamically varying the box constraint *C* and the RBF shape parameter $\sigma$ on the classification of IED events using SVM with an RBF kernel; a) Classification accuracy under combined geographic and temporal constraints using varying *C* and $\sigma$; b) Mean order of *C* at various window sizes; c) Mean order of $\sigma$ at various window sizes.**

PCA components is constrained to be ≤ 95% of total variance and the maximum *p*-value of the stepwise-selected features is constrained to 0.05. Results found using hyper-parameters generated using grid searched are marked with a 'G' in the legend. Results found using an estimated σ are marked in the legend with an 'E'. At each window size, geographic and temporal filters were applied to the dataset. Three-way cross-validation was used on the resulting subset to produce the data used in the Figure 50.a. Figures 50.b and Figure 50.c plot the change in mean hyper-parameter across geotemporal windows.

As shown in Figure 50.a, grid search outperforms $\sigma$ estimation at every window size (labeled as NDR G and PCA G). Interestingly, the best classification performance is seen in the smallest geographic windows, using NDR and PCA, with error rates approaching 20%. Figure 50.b shows that typical mean values of $C$ tend to be small. Mean values of σ tend to show little change with geographic window in Figure 50.c. The results for DF events are similar and can be found in Appendix D.

Figure 51 examines the impact and relationships of $C$ and $\sigma$ on classification error and the relationship between the $C$ and $\sigma$. Figure 51.a and Figure 51.b examine the impact of the order of $\sigma$ of classification error using grid search and estimation, respectively. In both cases, the error rate seems to be roughly constant across all values of $\sigma$. Figure 51.c and 51.d examine $C$ in a similar manner. As with $\sigma$, there seems to be little change in classification error rate across all values $C$ for both grid search and estimation.

Figure 51.e and 51.f examine the relationship between $C$ and $\sigma$. In Figure 51.e, there is a positive relationship, where increasing values of $C$ tend to be associated with

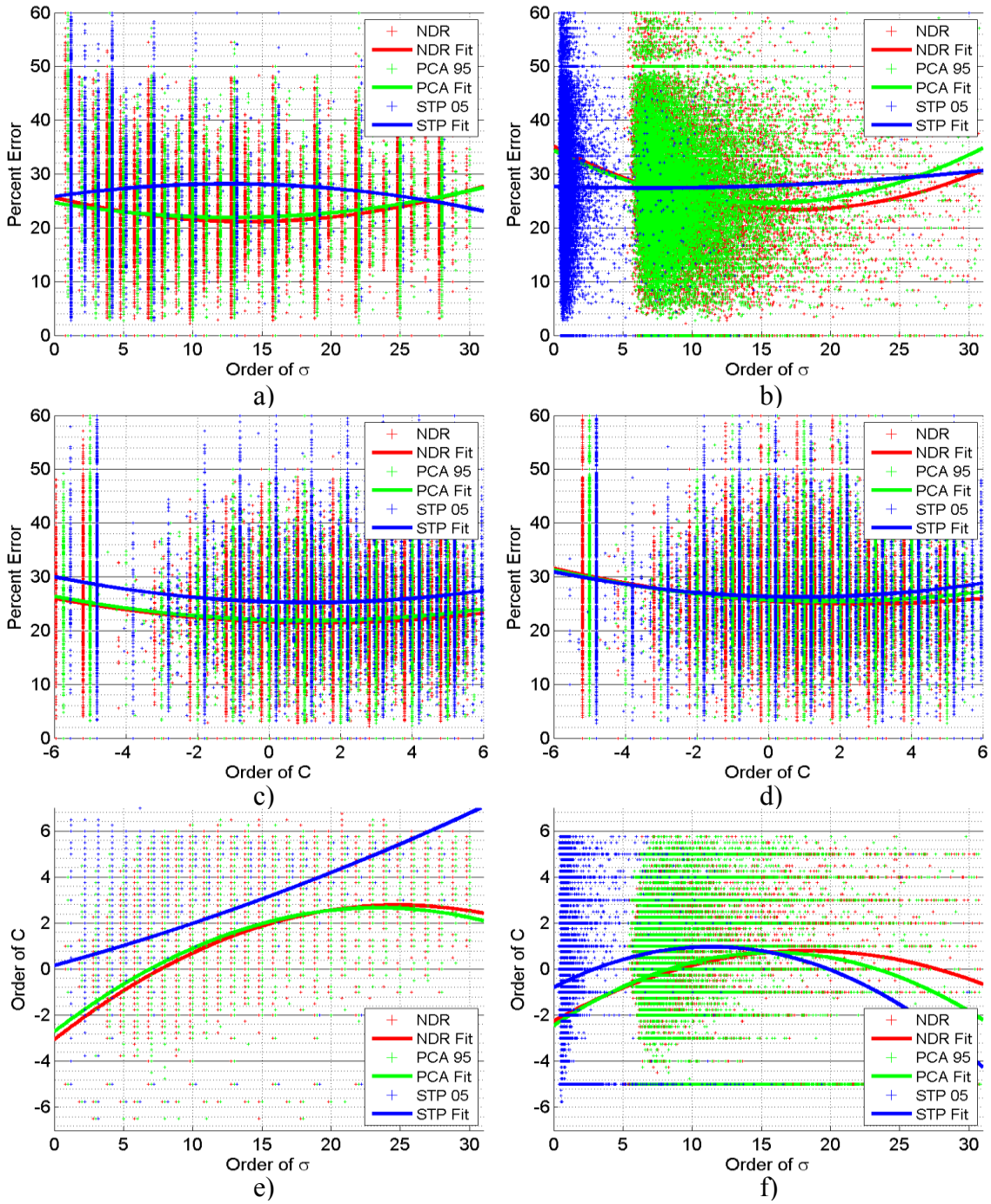



**Figure 51. Varying σ and C for IED events, a) Impact of the order of σ found using grid search on classification error; b) Impact of the order of σ found using estimation on classification error; c) Impact of the order of C found using grid search on classification error; d) Impact of the order of C found using estimation on classification error; e) Relationship between C and σ found using grid search; f) Relationship between C and σ found using estimation.**

increasing values of $\sigma$ until a general leveling out at values of $\sigma$ exceeding 20. The trend is somewhat similar in Figure 51.f for $C$ and $\sigma$ found using estimation, albeit with a less positive trend early and a decrease after values of $\sigma$ exceeding 20. The results for DF events are similar and can be found in Appendix D.

Figure 52 examines the impact of sample size on the order of $\sigma$ and $C$. For both grid search and estimation, the values of $C$ tend to increase with sample size. On the other hand, values of $\sigma$ climb slightly when grid search is used and remain roughly constant when estimation is used. The results for DF events are similar and can be found in Appendix D.

*Summary of Parameter and Hyperparameter Estimation*

Parameter and hyperparameter selection is an effective way to improve classification accuracy for both SVM and kNN. In the case of SVM, an RBF kernel with hyperparameters selected by grid search produced classification errors on the order to 20%. The lowest error occurred at the smallest windows and the smallest sample sizes, indicating that this combination may be the best overall in cases where the conflict event data is sparse. kNN also showed improvement by dynamically estimating $k$. Classification error of approximately 25% was consistent across all window sizes using this method.

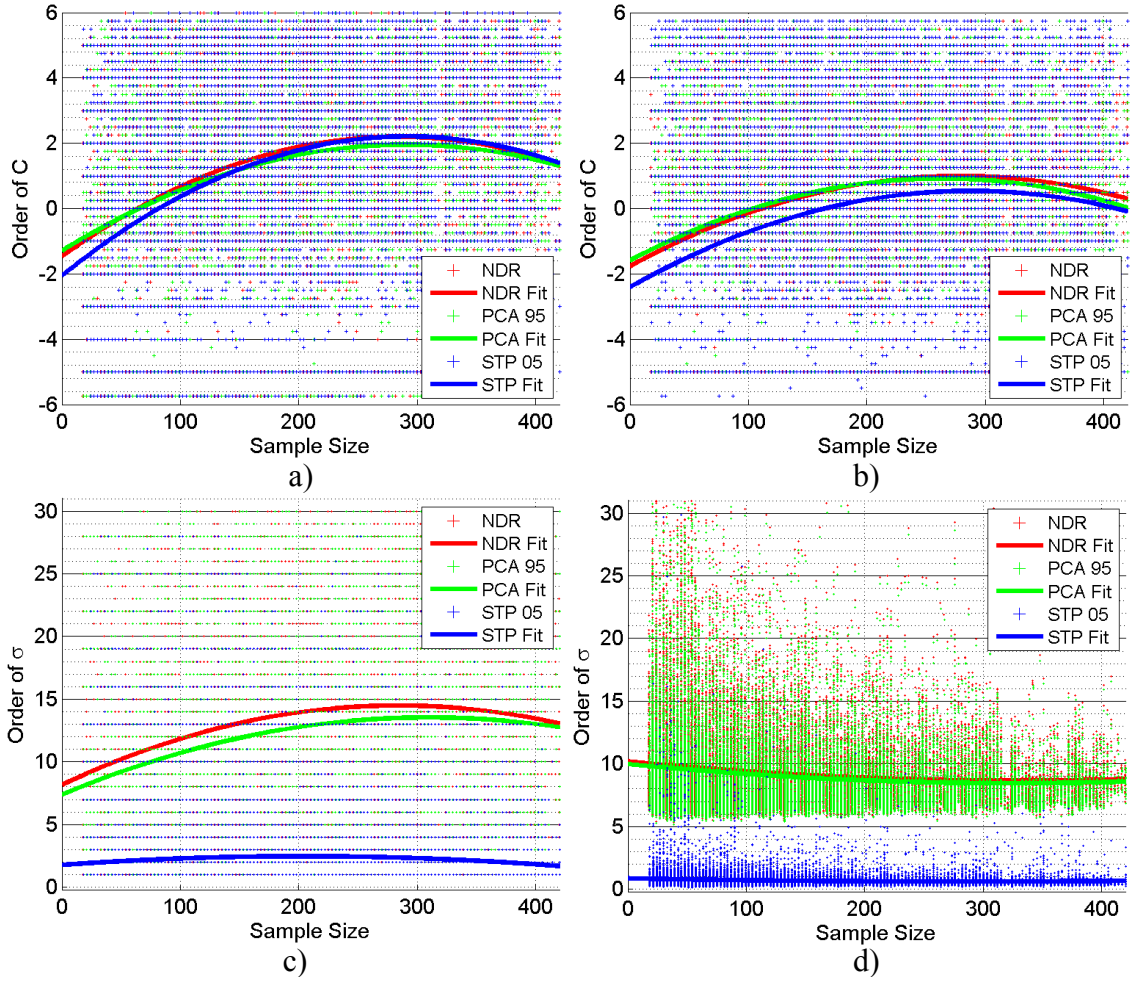



**Figure 52. For IED events, a) Impact of sample size on the order of *C* found using grid search; b) Impact of sample size on the order of *C* found using estimation; c) Impact of sample size on the order of *σ* found using grid search; d) Impact of sample size on the order of *σ* found using estimation.**

### A Cost Function for Estimating the Impact of Misclassification

Although overall error is an important metric for assessing the suitability of classification algorithms, it is not the only consideration. The cost of wrongly classifying a site is an important factor that varies between event and non-event classes. Another related, important consideration is potential mislabeling in the non-event training data.

To a soldier relying on an algorithm like MECH to predict hazardous terrain, a misclassified non-event site –a point on the road mistakenly labeled as an event site– is a nuisance. Frequent misclassifications of this type train the soldier to ignore the warning but otherwise fail to do harm. On the other hand, a misclassified event site may inadvertently portray a lethal situation as innocuous. In this case, a location that has utility for placement of an IED or direct fire attack is not correctly classified, potentially leading that soldier to proceed with less caution in a very hazardous situation. Thus, a misclassified event site has a much higher cost, potentially in human lives, than a misclassified non-event site.

Incorrectly classified non-event locations also have a cost, although not as severe. While misclassified events are always wrong, a "misclassified" non-event site might actually be classified correctly. This is because the true classification of a non-event site may vary from its classification in the dataset used in this research. The non-event class in the current dataset is composed of locations along roads in Afghanistan that are at least 250 meters from any known (to this research) IED or DF events. However, since the pool of known conflict events only covers 19 months in 2011-2012, it can be argued that the non-event class is actually a mix of three subclasses:

- non-event locations that will never be useful for a conflict event;
- locations that are useful for a conflict event but have not been used yet; and
- locations that have been used for a conflict event on a date outside this dataset.

Two of these three "non-event" subclasses should actually be labeled as "event" class. Effectively, this means that some fraction of the non-event class is actually mislabeled.

Given that a misclassified event is a definite error with potentially lethal consequences while the impact of a misclassified non-event is typically annoying and may actually be correct, it becomes important to develop a metric that weights the contribution of each type of misclassification. Equation 65 offers a weighted cost metric that weights and combines event and non-event errors:

$$Cost_W = \left(\frac{100 * |Misclass\ Evts|}{|Evts|}\right)^2 + C\left(\frac{100 * |Misclass\ NonEvts|}{|NonEvts|}\right) \qquad (65)$$

In common terms this can also be described as

$$Cost_W = (100 * (1 - precision))^2 + C(100 * negative\ predictive\ value)$$

In the equation, the percent of misclassified events is squared and then added to a weighted measure that consists of the percent of misclassified non-events scaled by a constant *C*. Note that the error measures are calculated independently for each class. The squared term causes $Cost_W$ to grow exponentially as event error increases. The growth of $Cost_W$ attributed to non-event error increases linearly.

The constant *C* defines the slope of the line describing non-event error and describes the x-axis value where this line and the curve describing event error intersect. For error values greater than *C*, misclassified events will contribute more to $Cost_W$ than misclassified non-events. Thus, *C* can be used to tune the sensitivity of the weighting algorithm to increased error. Low values of *C* cause the algorithm to weight event error more heavily than non-event error.

Figure 53.a examines the impact of the choice of *C* on the growth of $Cost_W$. In the figure, the event error term follows an exponential curve while the non-event error

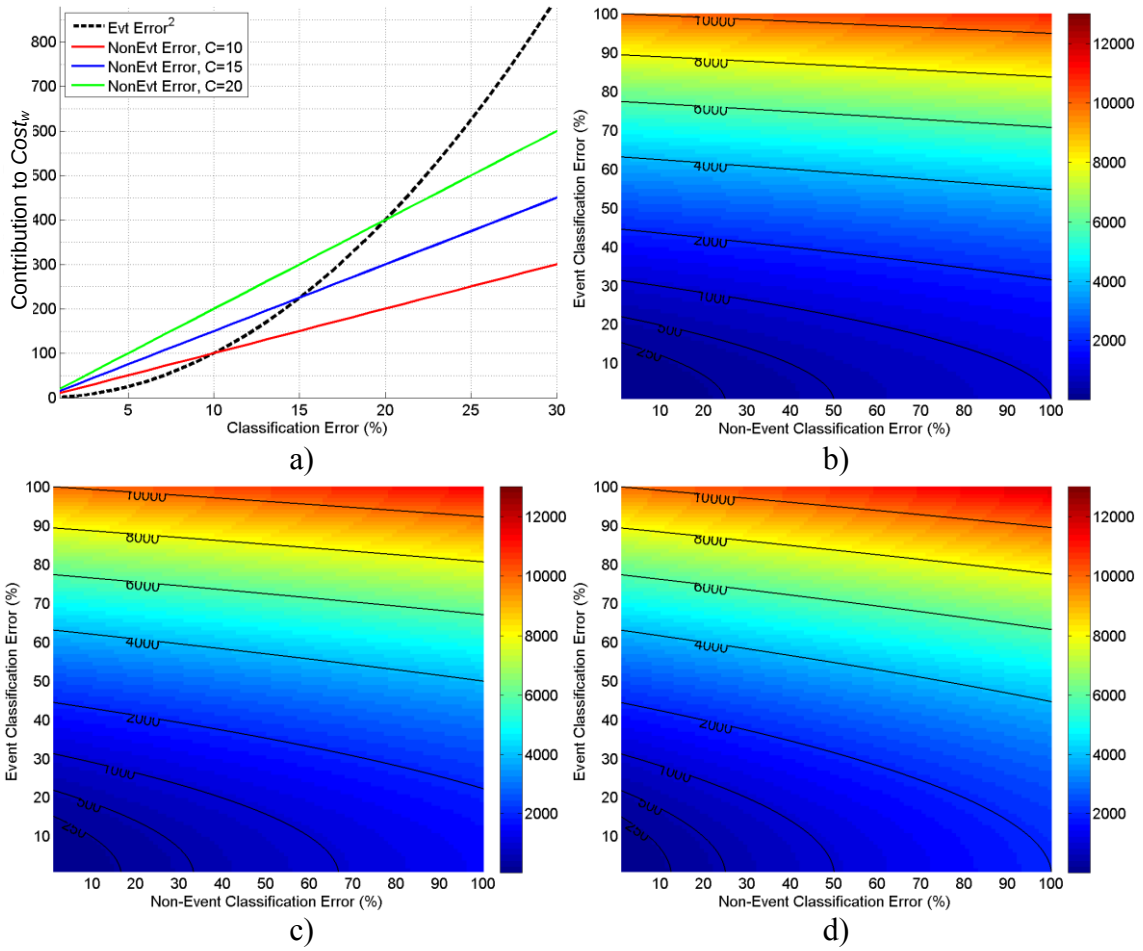



**Figure 53. Weighted cost incorporates both event and non-event error. a) Comparison of choices for *C*; b) $Cost_w$ when $C = 10$; c) $Cost_w$ when $C = 15$; d) $Cost_w$ when $C = 20$.**

term is linear, with the slope determined by *C*. Figures 53.b-d display the outcome of the weighting algorithm for event and non-event error values between 1% and 100% with *C* equal to 10, 15 and 20, respectively. Contour lines are added ease visual interpretation.

Given the potential human impact of mislabeling an event site, in the following analysis we set *C* to a value of 10. As shown in Figure 53.a, this means that the value of the non-event error component will tend to climb slowly as error increases. For $Cost_W$,

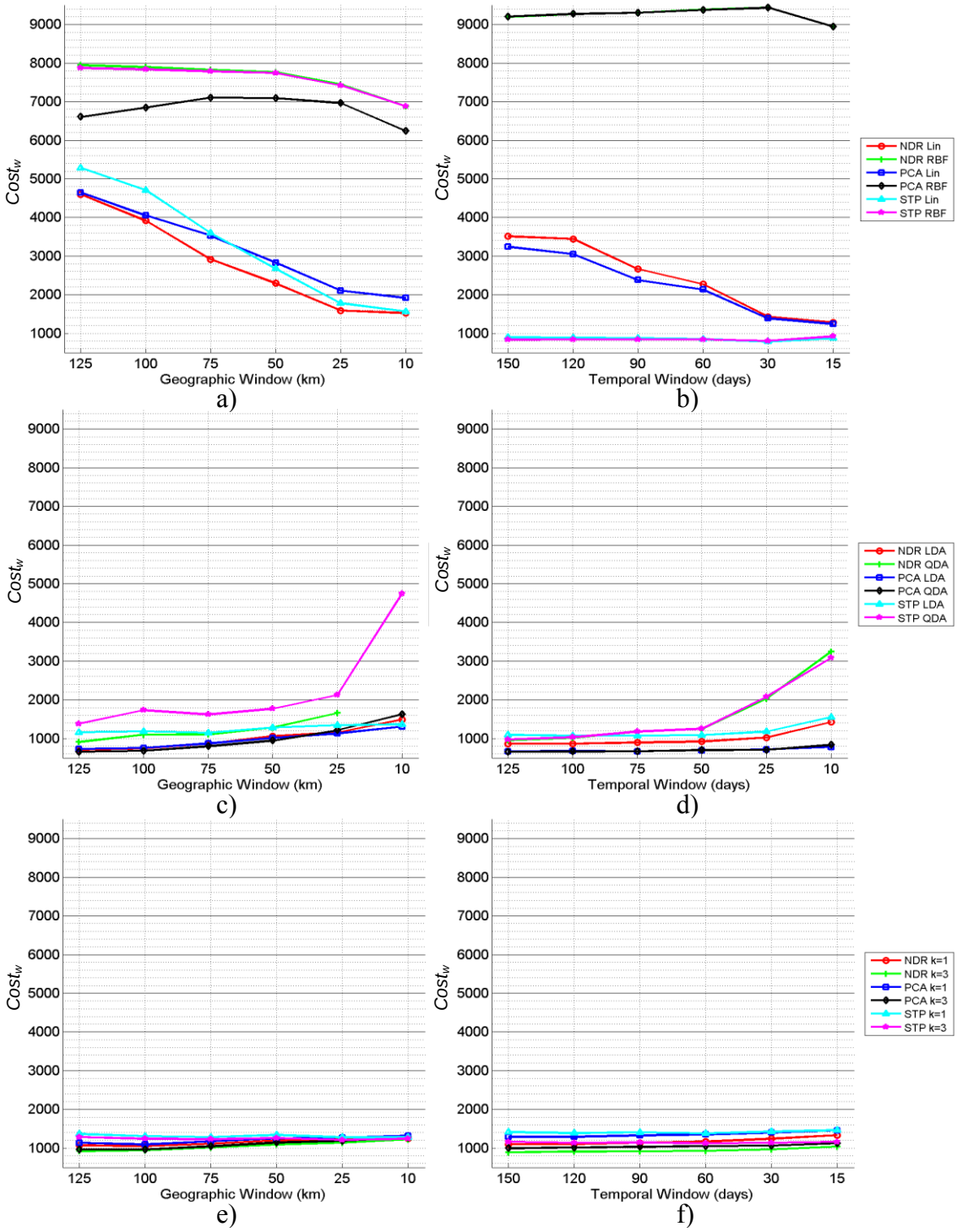



**Figure 54. Weighted cost from learning machines using default parameters; a) IED SVM with geographic constraints; b) IED SVM with temporal constraints; c) IED DA with geographic constraints; d) IED DA with temporal constraints; e) IED kNN with geographic constraints; f) IED kNN with temporal constraints.**

this emphasizes the contribution of event error and limits the impact of non-event error. The range of possible $Cost_W$ values for $C = 10$ is shown in Figure 53.b.

Figure 54 shows the weighted cost for the IED event class as classified by SVM, DA and kNN using default parameters with separate geographic and temporal constraints. The x-axis in each plot is the geographic or temporal window size. The y-axis is the weighted cost $Cost_W$. A constant scale is used across all plots in order to facilitate visual comparisons.

The behavior of different classification methods varies. The SVM classifiers using default parameters tend to perform poorly at large windows and large sample sizes where almost all of the $Cost_W$ comes from the event class. Within the SVM classifiers, the RBF kernels using default parameters tend to perform worst and the linear kernels show improved accuracy at smaller windows, both geographic and temporal. The DA classifiers, on the other hand, perform best at larger window sizes but tend to do slightly worse in the smaller windows. It is worth noting that DA with default parameters outperforms SVM at almost every window size. However, DA is unable to handle the small samples found using the smaller windows. kNN maintains a roughly constant error ratio across all window sizes.

The results for DF events are similar and can be found in Appendix D.

Figure 55 examines the weighted cost for the IED event class as classified by SVM, DA and kNN using optimized parameters with combined geographic and temporal constraints. The x-axis in each plot is the geotemporal window size. The y-axis is weighted cost. In the figure, weighted cost climbs as the geotemporal windows shrink.

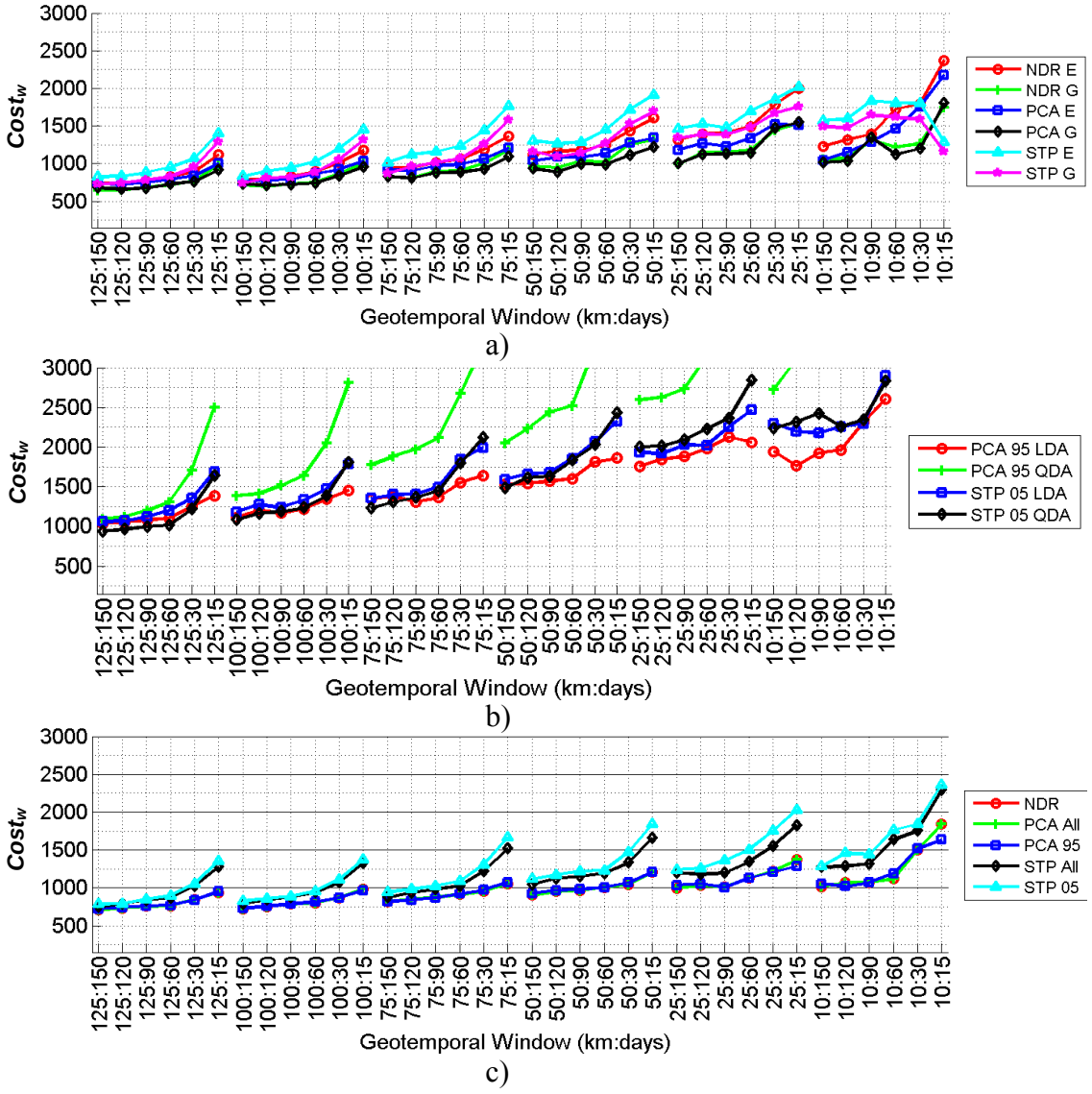



**Figure 55. Weighted cost from learning machines using optimal parameters on IED event data; a) SVM using an RBF kernel with geotemporal constraints; b) DA with geotemporal constraints; c) kNN with geotemporal constraints.**

SVM using NDR and PCA with parameters found by grid search performs best. DA and kNN show similar trends with the best overall performance coming from kNN. DA performs particularly poor at small window sizes. Overall, SVM using geotemporal

windows offers the best performance although the improvement over DA and kNN using separate geographic and temporal windows is relatively small.

The results for DF events are similar and can be found in Appendix D.

In summary, the cost of $Cost_W$ is a weighted cost metric that is produced by an algorithm that variably weighting event and non-event errors. Figures 54 and 55 show how $Cost_W$ with a $C$ parameter of 10 can be used to identify the best-performing combinations of learning algorithm, parameter optimizations and window types and sizes. As evidenced here, geotemporal windows seem to offer the best classification performance across all classifiers. Among classifiers, DA tends to perform poorly and classification accuracy decreases with the sample size. SVM slightly outperforms kNN and both classifiers seem to perform relatively well, even at small sample sizes.

In the following sections, $Cost_W$ will be used to assess the performance of several experiments. These experiments include dynamic geographic constraints, unbalanced classes, a fresh look at subset selection, and combinations of classifiers.

**Dynamic Geographic Constraints**

The best performing algorithms at this point are:

- SVM RBF kernel using PCA with hyperparameters estimated by grid search; and
- kNN using PCA with *k* estimated dynamically.

The training data used by these classifiers is normalized with location and scaling factors estimated from the data itself. Both classifiers use hyperparameters determined dynamically from the training set. However, up to this point, fixed geographic and temporal windows have been used to constrain the data used for training and testing the

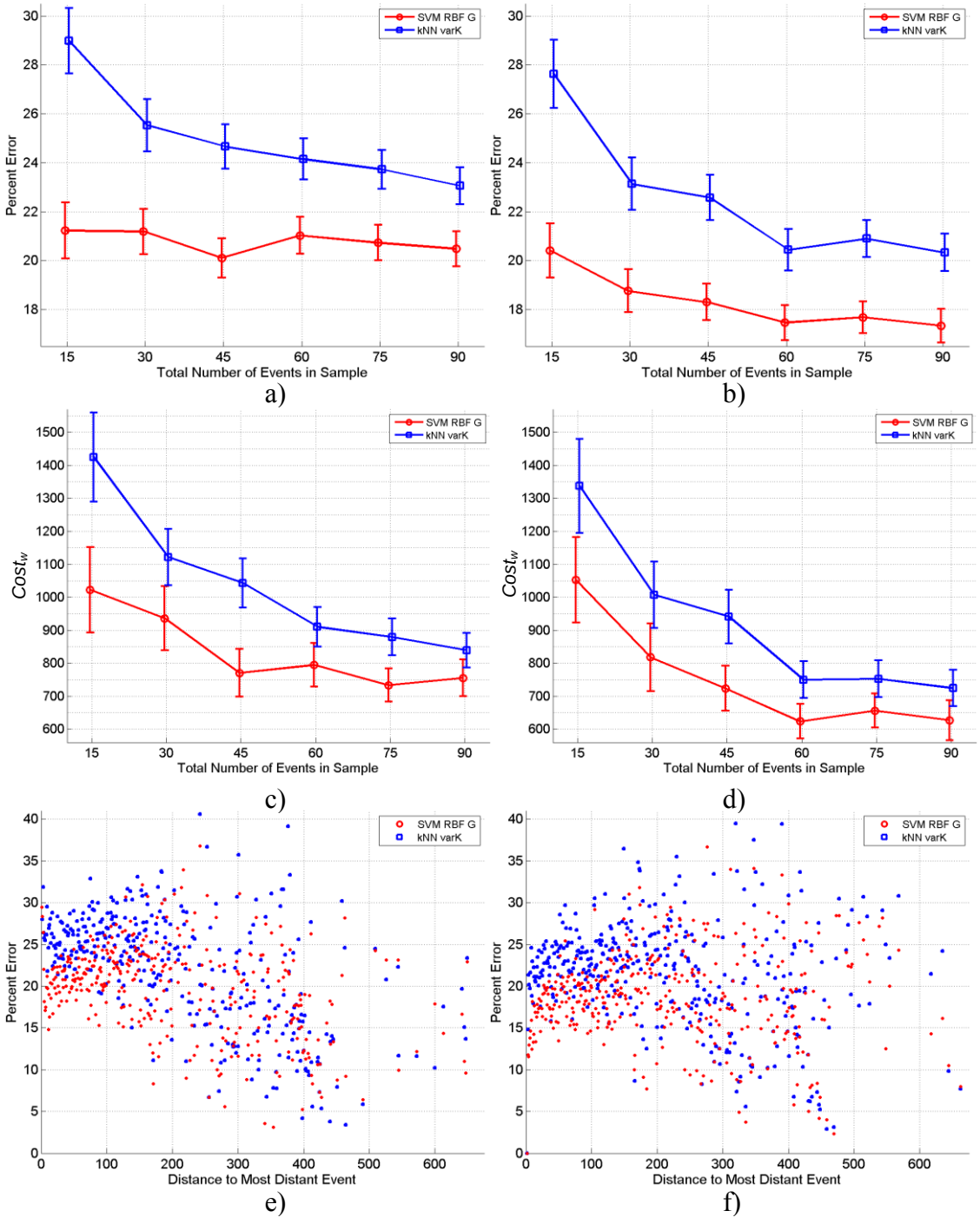



**Figure 56. Classification error using fixed sample sizes and dynamic geographic constraints; a) IED classification error; b) DF classification error; c) IED weighted cost; d) DF weighted cost; e) impact of geographic radius on classification error for IED; f) impact of geographic radius on DFclassification error.**

classifiers. The fixed geographic windows, in particular, may cause increased classification error because they fail to consider the distribution of events within the window. In many cases, there may be enough events to support the classifier's training within a geographic radius that is smaller than the fixed window. Since road points are selected from the entire window, it is likely that some of these road points are selected randomly from locations that are far away from any event. With larger windows, this may mean that road points are being selected from terrain that is very different than the terrain surrounding the events in the window. It may also mean that the fraction of total road points selected is smaller than necessary.

Given the performance of these two classifiers at relatively small window sizes, it seems possible that dynamic assignment of geographic radius may improve classification accuracy. In the following figures, the size of the event class is constrained to a set of fixed values: $n$ = [15 30 45 60 75 90]. The closest $n$ events to a selected center are selected and the distance from the center to the most distant member of $n$ is used as the geographic constraint. Road points are selected using this geographic constraint.

Figure 56 examines the impact of using a constant sample size to constrain the geographic window. In the experiment, the training sets were produced by first temporally constraining the data to a fixed window of 120 days before each *ST* event. From this temporally constrained set, the 15-90 events geographically nearest to the center (*ST* location) were selected. The distance of the most distant event from the center (*ST* location) was used as the radius of the geographic window and non-event points ($RD_{250}$) were selected from within this window. The test set was constructed by

temporally constraining events to a fixed window of 60 days after each *ST* event and geographically constraining events to the same radius used to select training data. Non-events points for the test were also selected using the same geographic constraint as the training set. Note that no cross-validation was used. Figure 56.a shows that the classification error rate using IED events approaches 20 % when using SVM, regardless of sample size. DF events produce even lower error rates in Figure 56.b. The weighted cost is shown in Figure 56.c and Figure 56.d. For both IED and DF events, particularly in the case of SVM, weighted cost decreases as sample size increases. For SVM, and to a lesser extent kNN, the decrease tends to level out around a sample size of 60. Across all plots in Figure 56, the best classification performance is produced by SVM. This performance advantage seems to be particularly high at the smallest geographic radii (Figure 56.e-f).

The consistent performance of SVM at small samples sizes using a dynamic geographic constraint is an interesting outcome. The results shown here are generally better and definitely more consistent than those seen using fixed windows.

**The Impact of Unbalanced Classes**

So far, all of the experiments shown in the research have used balanced classes. In other words, the number of events and the number of non-events were equal in the training and test sets. However, when large geographic and temporal constraints are used, the number of non-event points tends to outnumber events by a factor of 10,000 or more. Similarly, the number of non-events points selected may be very small in relation to the total number of non-event points found in the window. In comparison, ALL events

found in the window end up in either the training or set sets. In this case, it can be argued that the selection of balanced classes causes the event class to be over-represented in the training and test sets. To compensate, it is possible to intentionally create unbalanced training and test sets. However, this violates an often un-noticed assumption for many classifiers that the classes are equally represented. The resulting problems, often described as between-class imbalance, tends to grow worse as the imbalance gets more pronounced, potentially impacting the way in which *k* is chosen for kNN [146] or requiring sophisticated sampling methods to compensate [148].

Figure 57 examines the impact of unbalanced classes on predictive analysis using SVM and kNN. In the experiment, for each point in *ST*, the nearest 60 events are chosen for the training sample as described in the previous section. Six different sets of non-event points are chosen ranging in size from 60 to 360, in 60-point increments. The resulting event:non-event ratios go from 1:1 to 1:6. Six samples are produced from each entry in *ST* and used for training and testing with SVM and kNN. Note that no cross-validation was used.

In Figure 56.a and Figure 56.b, accuracy apparently improves, with classification error trending as low as 10% for both IED and DF events. However, this low error rate hides a serious problem: the misclassification rate of events increases greatly as the class imbalance grows. Figure 56.c and 56.d illustrate the problem and show that, at the largest imbalance, weighted cost is approximately six times larger than with balanced classes due to the large number of misclassified events. The lowest weighted cost occur

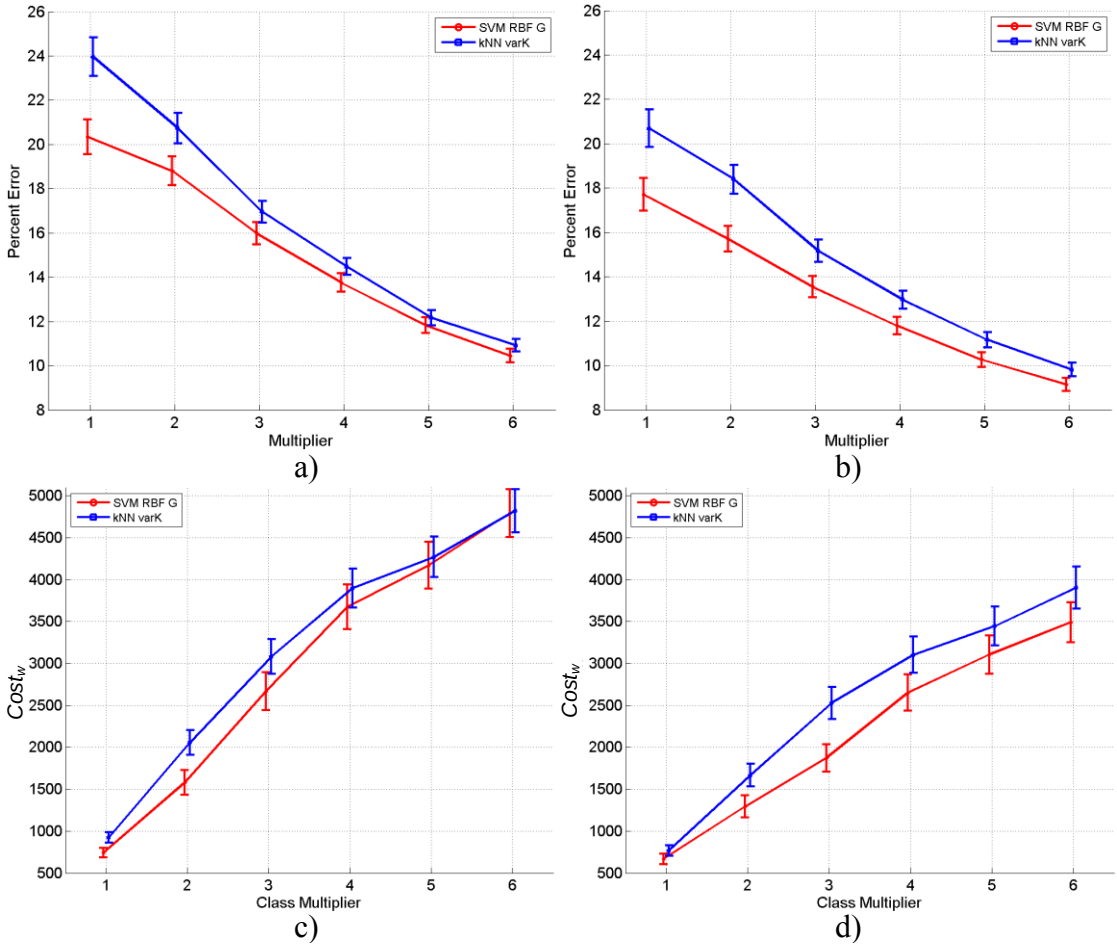



**Figure 57. The impact of unbalanced classes; a) IED classification error; b) DF classification error; c) IED event classification error ratio; d) DF event classification error ratio.**

when the classes are balanced. Similar findings are reported by Wei, et al in [149].

### Revisiting Feature Selection: Experts and Resource-constrained Environments

The best classifiers examined in this research produce mean classification errors on the order of 20% using a combination of dynamic techniques to choose training sets, normalize data, select features and classify. These classifiers rely only on information contained in the training sets to make classification decisions. Automated feature

selection using PCA and STP has been used extensively throughout this research. However, other approaches exist including feature selection by experts and feature selection for computationally constrained environments.

*Feature Selection Using Human and Automated Experts*

Recent conflicts in Iraq and Afghanistan have produced a generation of military personnel trained to detect and react to attacks. Many of these soldiers, sailors and Marines have convoyed and patrolled extensively in Afghanistan and developed a "sixth sense" or intuition about attack sites. Two of these experts, an Army Ranger, GP, and an Army Special Forces soldier, TN, were asked to identify features and distances that described a likely conflict event site. In their responses, both experts emphasized the importance of field of view for the attackers. TN focused on the importance of communications (no current feature captures this information), cover, concealment and escape adjacency for attackers. He also discussed the importance of terrain that restricts target movement, (possibly captured when a very short radial is found). GP focused on the characteristics of terrain required by attackers to support different types of attack. Their responses were used to select features from the existing set of 77. SG (the author of this dissertation) is also a former soldier and selected a feature set as well:

- TN: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, maximum and median cumulative escape adjacency, maximum and median route visibility at 100 meters
- GP: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, elevation, slope convexity, texture, roughness

at 350 m., discrete shape complexity at 1000 m., distance to populated area with at least one person

- SG: Short, long and median radial, local openness, planimetric area, rugosity, sparse viewshed shape complexity, slope convexity, texture, elevation range at 350 meters, distance to populated area with at least 1000 people, max and median route visibility at 100 m., max cumulative escape adjacency at 250 meters

Next, an automated subset selection method or blind expert uses feature correlation as a discriminator. Pearson's correlation coefficient was used to calculate the dependence between event features and non-event features,

$$\rho_{X,Y} = \frac{covariance(X,Y)}{\sigma_X \sigma_Y} \tag{66}$$

, where X and Y are an event feature and a non-event feature, respectively, and σ is the standard deviation of the feature. Equation 66 was used to build a correlation matrix, *M*, for the 77 features used in this research,

$$M_{i,j} = \rho_{event_i, non-event_j} \ for \ i,j = 1, \dots, 77 \tag{67}$$

The sum of the absolute value of each column of *M* is inspected. The ten features with the lowest cumulative correlations are selected.

The experiment to test these different subset selection methods used fixed sets of 60 events and 60 non-events chosen using:

- fixed temporal training and test windows of 120 and 60 days, respectively;
- dynamic geographic windows, with a radii equal to the distance from the most distant event to the *ST* location.

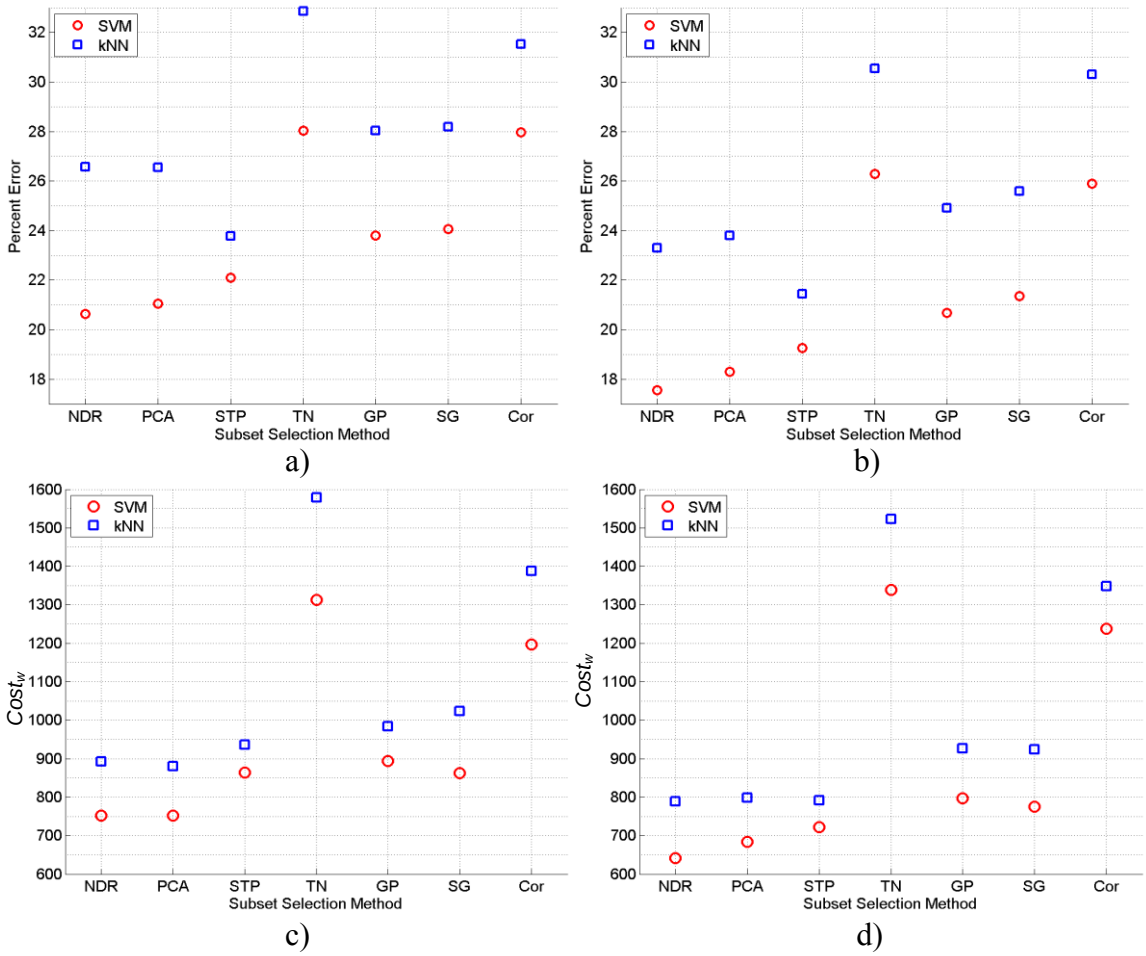



**Figure 58. Alternate methods for subset selection; a) IED classification error; b) DF classification error; c) IED event classification error ratio; d) DF event classification error ratio; e) IED overall error vs event classification ratio; f) DF overall error vs event classification ratio.**

- No cross-validation

For each event in *ST*, the resulting training and test sets were used to compare the classification accuracies the three experts (GP, TN, SG), the blind expert (COR), and the more conventional feature selection approaches of NDR, STP and PCA.

**Table 5. Event classification error, from experts (% of total sample).**

| | IED (%). | | DF (%). | |
|---|---|---|---|---|
| | SVM | kNN | SVM | kNN |
| NDR | 9.7 | 10.1 | 8.9 | 9.5 |
| PCA | 9.7 | 10.0 | 9.1 | 9.5 |
| STP | 10.2 | 11.4 | 9.1 | 10.1 |
| TN | 14.3 | 16.3 | 14.4 | 15.8 |
| GP | 10.3 | 10.8 | 9.7 | 10.5 |
| SG | 10.0 | 11.5 | 9.2 | 10.3 |
| Cor | 12.8 | 14.6 | 13.3 | 14.2 |

Figure 58 shows the results of the experiment where experts select the best features. Across the board, SVM produces lower overall classification error rates for both IED and DF. The same is true for weighted cost, although the difference SVM and kNN tends to be less pronounced. With the exception of two experts, all of the methods shown are at or below a weighted cost of 1000. Table 5 shows the actual event classification error rate, or the percent of the total sample that consists of misclassified events. As shown in the table, the best result for IED classification is SVM NDR and PCA where 9.7% of the total sample consists of misclassified events.

*Feature Selection for Resource-constrained Computing Environments*

Military personnel tend to operate in environments with severe and strict constraints on the size, weight and power (SWAP) of computing and communications devices. Computing devices suitable for battlefield use must be small and light enough to easily transport but rugged enough to handle combat stresses. Power is shared by communications, computing and surveillance, and power provisioning contributes

substantially to the overall load carried by the soldier. Thus, efficient use of computing and communications resources is critical.

Two common approaches to solving this problem include pre-computation of features and offloading computing to other devices. However, these solutions only shift the burden and may not actually solve any problem. For example, for many terrain-based features, it is possible to exhaustively compute the feature at small intervals across an entire battlefield. However, the information must be stored and retrieved. The computing device must have either large amounts of storage or frequent access to communications, both of which increase the power use and potentially increase the complexity of the computing device. Offloading computations incurs a similar communications cost. A difficult-to-predict, but nevertheless real, related problem is that modern battles are also fought in the electromagnetic spectrum. In today's warfare, communications may be intentionally disrupted or targeted and will not always be available.

Given these known power and communications constraints, the computational cost of feature collection is an important consideration. This cost varies and cost-informed feature collection can reduce power and enable effective computation on resource-constrained devices. In some cases, elevation for example, the feature is collected through a simple lookup. Features like elevation range and roughness involve operations on an $n \times n$ matrix but the size of the matrix is constrained and known beforehand, and the operations themselves are relatively simple. At the other extreme are features based on discrete viewshed, which requires a large number of relatively costly operations. Discrete shape complexity index is an example of this type of feature.

Cumulative escape adjacency-based features, built from combinations of many viewsheds, are even more costly. Thus, the actual features collected are an important consideration for computationally constrained computing.

Given the constraints on SWAP and the nature of the modern battlefield, we assume that:

- communications may be unavailable;
- power is constrained; and
- computing power and storage is limited.

Thus, solutions considered should involve only local, relatively inexpensive calculations on locally available data. In this research, we examine the impact of feature selection on computing requirements under these constraints and assess the tradeoff between accuracy and computing requirements. A total of seven feature sets are considered and compared with the now familiar solutions of NDR, PCA and STP.

The first five features sets, Rad4 through Rad64, are built using seven features collected from sparse viewshed: short radial, long radial, mean radial, local openness, planimetric area, rugosity and shape complexity. Five different features sets are created by collecting these seven features at five different resolutions: 4, 8, 16, 32, and 64 radials.

The sixth feature set, Local, is built using 9 features that can be calculated by loading a single $n \times n$ matrix of elevations and one feature that requires a table lookup and distance measurement: elevation range, roughness, Rad64 and distance to nearest populated area with >1000 people.

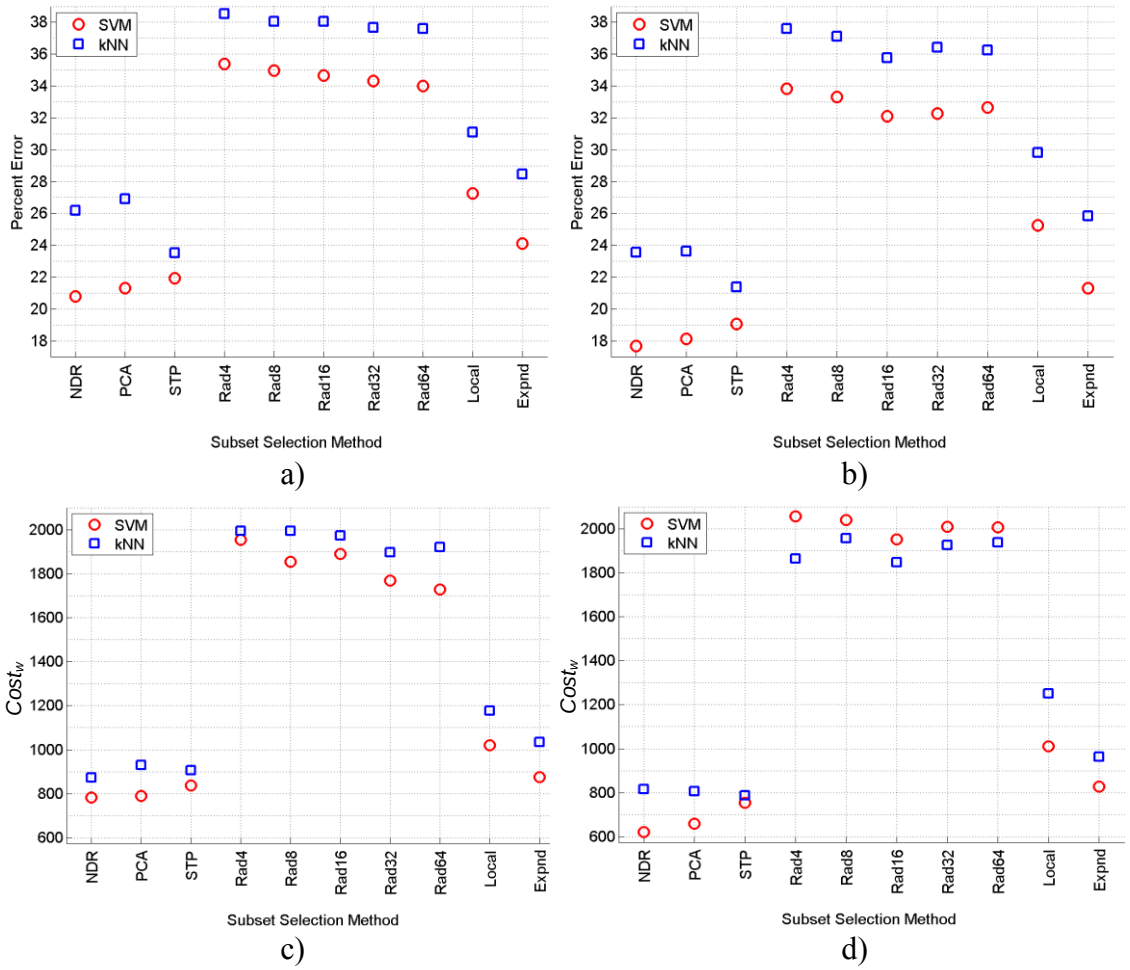



**Figure 59. Computationally inexpensive subsets; a) IED classification error; b) DF classification error; c) IED weighted cost; d) DF weighted cost.**

The seventh feature set, Expnd, is constructed similarly using 15 features, augmenting the Local set with three geomorphometric features commonly used to classify terrain -slope, convexity, and texture- and the distance to nearest populated areas with >1, and >10000 people.

Figure 59 examines the classification error produced by the seven resource-constrained feature sets and compares it to NDR, PCA and STP. In the figures, Rad*

**Table 6. Event classification error, from resource-aware feature selection (% of total sample).**

| | IED (%). | | DF (%). | |
|---|---|---|---|---|
| | SVM | kNN | SVM | kNN |
| NDR | 10.0 | 9.8 | 8.8 | 9.6 |
| PCA | 10.0 | 10.2 | 8.9 | 9.6 |
| STP | 10.0 | 11.0 | 9.5 | 10.0 |
| Rad4 | 17.6 | 18.7 | 18.5 | 18.0 |
| Rad8 | 17.0 | 18.8 | 18.6 | 18.7 |
| Rad16 | 17.4 | 18.8 | 18.4 | 18.2 |
| Rad32 | 16.9 | 18.2 | 18.7 | 18.6 |
| Rad64 | 16.7 | 18.3 | 18.9 | 18.8 |
| Local | 10.7 | 12.5 | 11.1 | 13.1 |
| Expnd | 10.2 | 11.3 | 10.1 | 10.8 |

features sets tend to perform poorly with high error and weighted cost. Local performs better than Rad* but still fails to reach the accuracies of NDR, PCA and STP. Interestingly, Expnd performs fairly well with only 17 features and no additional feature selection. Although overall error is 2-3 percent higher than than NDR and PCA, the weighted cost for Expnd is not much larger than these other feature selection algorithms. This seems to indicate that it may be possible to perform the classification task using a reduced set of selected, computationally inexpensive features without incurring a high penalty.

Table 6 shows the actual event classification error rate, or the percent of the total sample that consists of misclassified events. As shown in the table, the best result for IED classification is SVM NDR and PCA where 10% of the total sample consists of misclassified events. Expnd is close with an event classification error rate of only 10.2%.

## Combining Classifiers

The best individual classifiers used in this research produce classification errors near 20%. Both SVM and kNN perform well although SVM does better with small training sets. For both SVM and kNN there seems to be a tradeoff: decreased overall classification error comes at the cost of increased event classification error. Since a misclassified event site has a large real-world cost —an IED explodes or an ambush is not anticipated— we prefer to find classifiers that reduce misclassified events while keeping total classification error as small as possible. Ensemble-based classifiers offer a potential way to achieve this goal.

Ensemble-based classifiers build predictive models by combining the output of multiple classifiers [150]. In this research we examine the effectiveness of ensemble classifiers using a majority vote rule, a cost-sensitive rule, and stacking. The ensembles are constructed using a variety of common and mixed base classifiers with both overlapping and disjoint feature sets.

Hastie [81] notes that a good ensemble requires a collection of base classifiers that cover the input and output space and are sufficiently different from one another. While he does not define 'sufficiently different', other authors propose strategies using disjoint feature subsets to ensure this difference [151], [152]. The MECH model offers a clear way to produce these disjoint sets: the set of Emplacement features and the set of Monitor/Control features. As a reminder, Emplacement features describe the event site itself and the area immediately adjacent to it. These features include geomorphometric measures of the terrain and LOS-based assessments of local terrain. Monitor/Control

features attempt to capture the embedding of the Event site into the surrounding terrain through line-of-sight analysis and large-scale geomorphometric measures.

Ensemble-based classifiers make classification decisions by combining the output of multiple individual classifiers according to some rule or algorithm. In this research we examine three cases: majority vote rule, cost-sensitive rule, and stacking using SVM and kNN algorithms. The majority vote rule assigns a class label by simply counting the number of 'votes' for each class from individual classifiers. Ensemble-based classifier *C* is composed of the output of multiple individual classifiers. Let $c_i$ be the output of an individual classifier containing *n* classifications and let $c_{ij}$ be its *j*-th output. Then, the classification output of ensemble-based classifier *C* is assigned by a majority vote of its individual classifiers,

$$C_{i,j} = \begin{cases} event \mid \sum_i \left( c_{i,j} \equiv event \right) \geq \sum_i \left( c_{i,j} \equiv nonev \right) \\ nonev \mid \sum_i \left( c_{i,j} \equiv nonev \right) > \sum_i \left( c_{i,j} \equiv event \right) \end{cases} for\ j = 1, \ldots, n \qquad (69)$$

The cost-sensitive rule takes into account the real-world cost of a misclassification and prefers to misclassify non-events. If any member of the ensemble classifies a location as an event, then the ensemble classifies it as an event,

$$C_{i,j} = \begin{cases} event & \mid \sum_i \left( c_{i,j} \equiv event \right) > 0 \\ nonev & \mid \sum_i \left( c_{i,j} \equiv event \right) \equiv 0 \end{cases} \quad for\ j = 1, \ldots, n \qquad (70)$$

Note that when the ensemble consists of two classifiers, both of these rules produce the same result.

Voting algorithms simply combine the output of existing classification outcomes according to some rule. Stacking, on the other hand, takes the same combinations of classification outcomes and uses them as input to a classification algorithm. Instead of simply counting outcomes, stacking combines the output of multiple base classifiers to form a new dataset. This new dataset, composed of binary classification algorithm results, is used to train a new classifier. In this research, several different base classifier combinations, or stacks, are used with SVM and kNN.

*Ensembles Constructed of Single Algorithm Base Classifiers*

Figure 60 examines the classification error using SVM and kNN of three ensembles using two different rules. In each ensemble, all classifiers use a common, single algorithm, either SVM or kNN. The outcome of the ensembles is compared to several individual classifiers, including:

- NDR, PCA and STP;
- Emp: a classifier that uses the Emplacement features from Table 4; and
- MC: a classifier that uses the Monitor/Control features from Table 4.

The ensembles include:

- NPS: the classification outcomes of {NDR, PCA, STP};
- All: the classification outcomes of {NDR, PCA, STP, Emp and MC}; and
- EMC: the classification outcomes of {Emp and MC}.

For the three ensemble classifiers, majority vote and cost-sensitive vote rules are used.

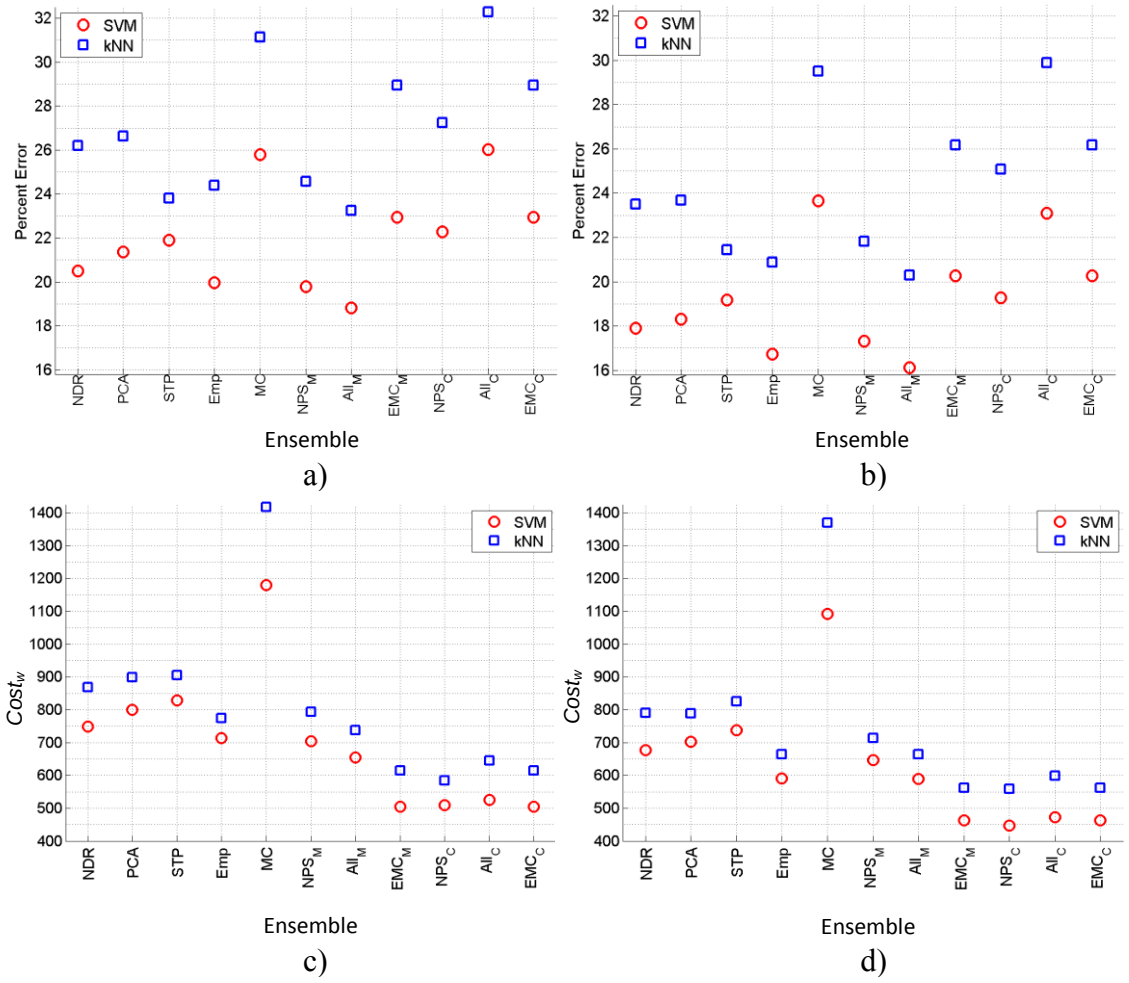



**Figure 60. The classification accuracy of single base classifier ensembles; a) IED classification error; b) DF classification error; c) IED weighted cost; d) DF weighted cost.**

-

Figures 60.a and 60.b examine the classification accuracy of the ensembles. Although $All_M$ performs slightly better than others, the difference is not pronounced. Figures 60.c and 60.d, however, do show a significant improvement in the weighted cost by the ensembles. All of the ensembles using the cost-sensitive rule show significant decrease in event classification error. Note that $EMC_M$ and $EMC_C$ show the same

performance because the two rules, majority and cost-sensitive, produce the same results when there are only two ensemble members.

Table 7 shows the actual event classification error rate. As shown in the table, combining classifiers can significantly reduce misclassified events. Although the best result for IED classification is $All_C$ at 2.1%, a close second is the MECH model-based EMC at 3.8%.

**Table 7. Event classification error, from single algorithm base classifiers (% of total sample).**

| | IED (%). | | DF (%). | |
|---|---|---|---|---|
| | SVM | kNN | SVM | kNN |
| NDR | 9.6 | 9.5 | 9.2 | 9.2 |
| PCA | 9.9 | 9.5 | 9.1 | 9.3 |
| STP | 10.1 | 11.1 | 9.1 | 10.1 |
| Emp | 9.0 | 9.1 | 8.5 | 8.2 |
| MC | 13.0 | 14.3 | 12.9 | 14.4 |
| $NPS_M$ | 9.1 | 8.8 | 8.7 | 8.6 |
| $All_M$ | 8.6 | 8.4 | 8.4 | 8.3 |
| $EMC_M$ | 3.8 | 3.7 | 4.5 | 3.9 |
| $NPS_C$ | 4.6 | 3.7 | 4.6 | 4.1 |
| $All_C$ | 2.1 | 1.6 | 2.5 | 1.9 |
| $EMC_C$ | 3.8 | 3.7 | 4.5 | 3.9 |

*Ensembles Constructed of Mixed Algorithm Base Classifiers*

The experiment shown in Figure 60 only addressed ensembles composed of outcomes using the same base classifier, KNN or SVM. However, ensembles composed of mixed algorithm base classifiers offer some improvement by increasing diversity [150]. Figure 61 shows the results of using ensembles with mixed algorithm base classifiers. In the figure, seven ensembles are examined:

- SK3: [ SVM-NDR, SVM-PCA, SVM-STP, kNN-NDR, kNN-PCA kNN-STP];
- SK-All: [SVM-NDR, SVM-PCA, SVM-STP, SVM-Emp, SVM-MC, kNN-NDR, kNN-PCA, kNN-STP, kNN-Emp, kNN-MC];
- SK-EMC: [SVM-Emp, SVM-MC, kNN-Emp, kNN-MC];
- SE-KEMC: [SVM-Emp, kNN-Emp, kNN-MC];
- SEMC-KE: [SVM-Emp, SVM-MC, kNN-Emp];
- SN-KS: [SVM-NDR, kNN-STP];
- SN-KSEMC: [SVM-NDR, kNN-STP, kNN-Emp, kNN-MC].

In Figure 61, the overall error rates show a slight improvement in some cases. However, the weighted cost remains on par with the best rates seen in Figure 60. The range of weighted cost values is relatively small, with most ensembles found between 500 and 600. The best overall classifiers for IED events by weighted cost are probably SK-EMC$_M$ and SN-KSEMC$_M$. The two ensembles have error rates near 20% and weighted cost around 510.

Table 8 shows that all of the mixed algorithm base classifiers perform fairly well. Of particular note is the performance of the mixed classifiers SE-KEMC and

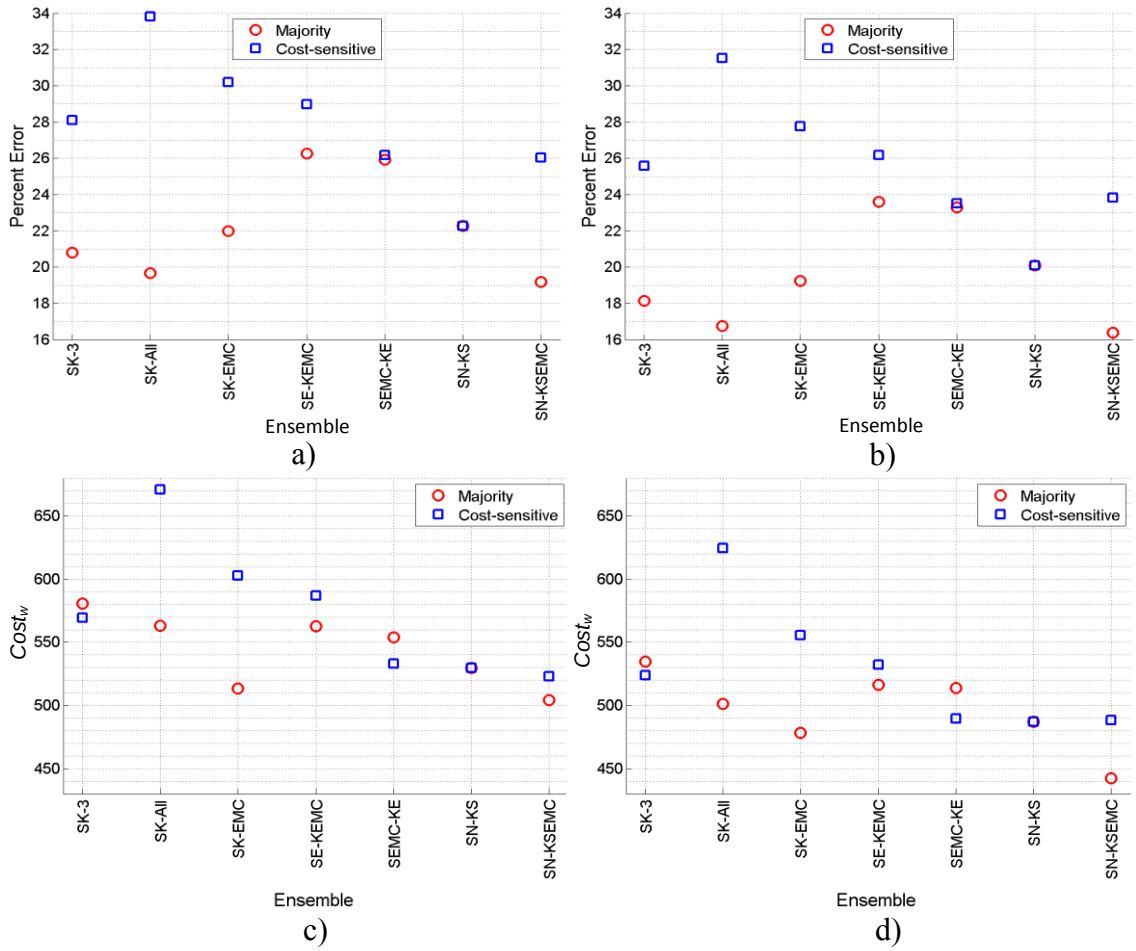



**Figure 61. The classification accuracy of mixed base classifier ensembles; a) IED classification error; b) DF classification error; c) IED event classification error ratio; d) DF event classification error ratio; e) IED overall error vs event classification ratio; f) DF overall error vs event classification ratio.**

SEMC-KE. Both of these MECH model-based classifiers perform well using both majority and cost-sensitive rules. This performance supports the MECH model concept of dividing analysis into two spaces: features collected from the conflict event Emplacement site and features collected from the conflict event Monitor/control area surrounding the Emplacement site.

**Table 8. Event classification error, from mixed algorithm base classifiers (% of total sample).**

| | IED (%) | | DF (%) | |
|---|---|---|---|---|
| | Majority | Cost-sensitive | Majority | Cost-sensitive |
| SK-3 | 6.8 | 2.4 | 6.9 | 2.7 |
| SK-All | 6.9 | 0.8 | 6.7 | 1.1 |
| SK-EMC | 4.8 | 1.7 | 5.2 | 2.1 |
| SE-KEMC | 3.7 | 2.5 | 4.0 | 2.7 |
| SEMC-KE | 3.6 | 2.4 | 3.9 | 2.9 |
| SN-KS | 5.3 | 5.3 | 5.2 | 5.2 |
| SN-KSEMC | 6.0 | 2.1 | 6.1 | 2.5 |

*Stacking*

The final experiment in this category examines the impact of stacking on classification error. Stacking, or stacked generalization, uses an ensemble to train a learning algorithm. The ensemble is created by combining the classification decisions of two or more individual classifiers into a new dataset. In this dataset, each feature is actually the outcome of an individual classifier that was trained on original data. This new, composite dataset is used to train a new classifier, which then makes a final classification decision or prediction. In this experiment, various combinations of base classifiers are used to create 13 different stacks. The stacks include the ensembles used in Figures 60-61.Each different stack is used to train and test SVM and kNN classifiers.

Figure 62 examines the impact of stacking. In the figure, the 13 ensembles are listed on the x-axis. The first six ensembles use subscripts to describe the algorithm used

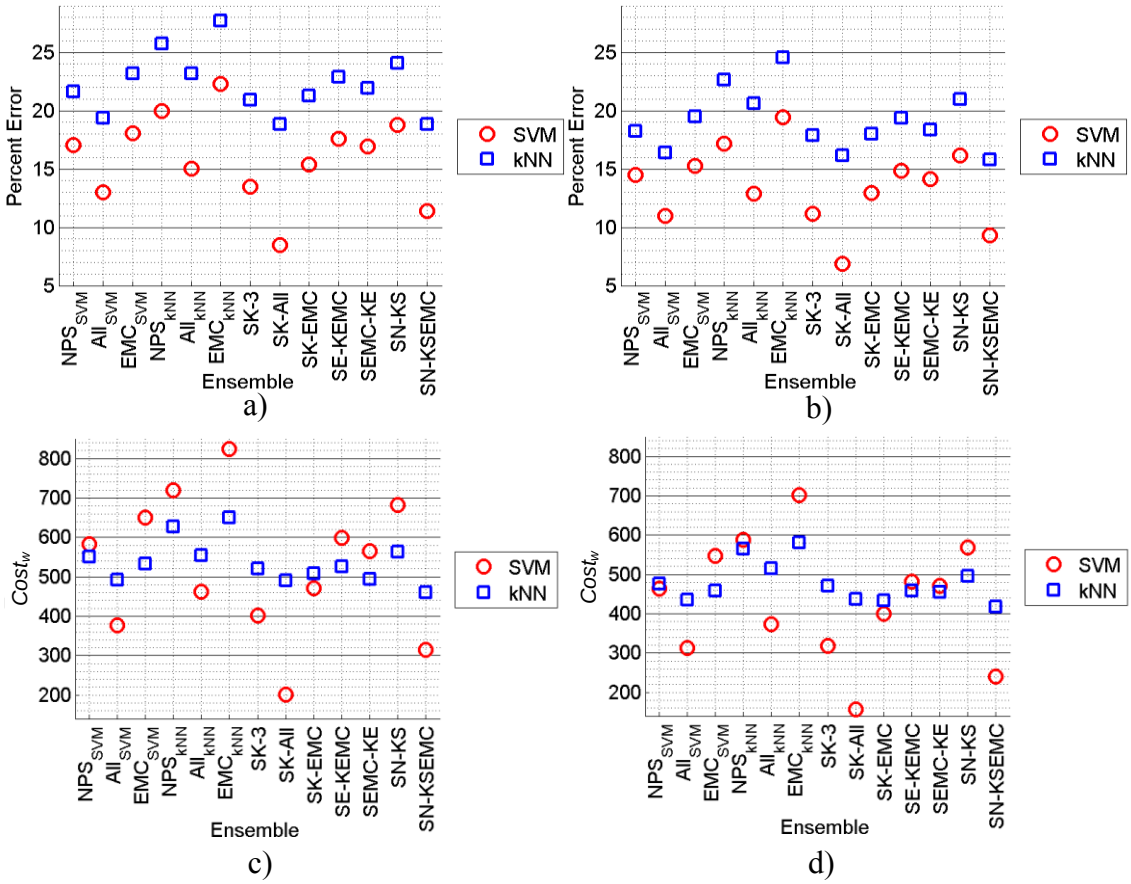



**Figure 62. The classification accuracy of stacking; a) IED classification error; b) DF classification error; c) IED weighted cost; d) DF weighted cost.**

by the original base classifiers. So, the $NPS_{SVM}$ ensemble is composed of three classification outcomes produced by using SVM on original data with dimensionality reduction schemes of NDR, PCA, and STP. In the figure, the overall classification error is reduced across all ensembles, especially for outcomes predicted using SVM. The improvement is particularly pronounced for SK-ALL using SVM, which has an overall error rate for IEDs of around 8.5 % and an event classification rate of 4.1%.

**Table 9. Event classification error, from stacking using SVM and kNN (% of total sample).**

| | IED (%). | | DF (%). | |
|---|---|---|---|---|
| | SVM | kNN | SVM | kNN |
| NPS-SVM | 8.1 | 7.0 | 7.5 | 7.1 |
| All-SVM | 5.9 | 6.5 | 5.8 | 6.6 |
| EMC-SVM | 8.7 | 4.7 | 8.5 | 5.5 |
| NPS-kNN | 9.0 | 6.8 | 8.4 | 6.7 |
| All-kNN | 6.6 | 6.0 | 6.2 | 6.2 |
| EMC-kNN | 9.9 | 5.4 | 9.1 | 5.5 |
| SK-3 | 5.9 | 6.4 | 5.6 | 6.4 |
| SK-All | 4.1 | 6.3 | 3.8 | 6.5 |
| SK-EMC | 7.2 | 5.3 | 6.8 | 5.6 |
| SE-KEMC | 8.0 | 5.5 | 7.5 | 5.8 |
| SEMC-KE | 8.0 | 5.5 | 7.6 | 6.0 |
| SN-KS | 8.7 | 5.4 | 8.1 | 5.3 |
| SN-KSEMC | 5.2 | 5.9 | 4.9 | 6.0 |

Table 9 shows that stacking does not perform as well as single algorithm and mixed algorithm base classifiers when event classification error is the primary concern. However, as shown in Figure 62, SK-$All_{SVM}$ performs better than any other classification scheme when total error or weighted cost are used to make the decision. The best performance comes with a relatively high computational cost: a dataset consisting of 60 events and 60 non-events is classified five times with SVM and five time with kNN. The ten sets of classification outcomes are used to create a new feature set that is then classified using SVM. The decision to use this relatively heavy-weight process will depend on the computational resources and time available to the user.

## Analysis of Event Classification Errors

The previous sections identify several classifiers that produce relatively low overall classification error while maintaining low event classification error. However, event classification error for the best classifiers is still on the order of 3-10% of the total sample. The following analysis examines event classification error using seven classifiers and examines their performance with regards to potentially strong influencing factors like terrain type, time, date and distance from populated areas. The seven classifiers are:

- Individual classifiers: SVM-NDA, SVM PCA, SVM-STP;
- MECH-based ensembles using a single base classifier: SVM-EMC, kNN-EMC;
- MECH-based ensembles using mixed base classifiers: SE-KEMC, SEMC-KE.

For the analysis, a dataset was assembled that contains the test outcomes individual test set samples for each of the seven classifiers using the 500 events in *ST* as seeds. The 500 test sets in the dataset collectively contain 22926 sampled IED events (DF: 21616). These events are sampled from a set of 2865 unique IED events (DF: 2216), so the average event appears multiple times. Individual IED event counts range from 1 to 28 (DF: 29) with a mean of 8.0 (DF: 9.8).

Some IED events are easier to correctly classify than others. In the dataset, a total of 808 unique events (representing 4745 samples) are always classified correctly by all seven classifiers (DF: 630, 5338). This analysis can expanded to include events that are always correctly classified by a majority of the seven classifiers. In this case, 1725

unique events (representing 12197 samples) are always correctly classified by four or more of the seven classifiers (DF: 1267, 12056).

Misclassified events can be examined similarly. A total of 22 events (representing 31 samples) are always misclassified by all seven classifiers (DF: 11, 20). Similarly, 53 events (representing 96 samples) are always misclassified by at least four of the seven classifiers (DF: 27, 74).

In the analysis that follows, we focus on events that are misclassified by a majority of the seven classifiers (four or more); and misclassified a majority of times that they appear. So, for example, a unique event that appears in 11 test sets will be included in this analysis if it is misclassified in at least 6 (of 11) samples by at least four (of 7) classifiers. Based on these criteria, the analysis will include 387 unique events (representing 1226 samples). These are compared to events that are classified correctly in a majority of samples by a majority of classifiers. Comparisons are global across the dataset

Figure 63 examines the effect of terrain type on the classification of events. In the figure, three histogram bars are associated with each terrain type. The first bar represents majority misclassified and majority correctly classified events in a stack, denoted by red and green segments, respectively. The middle bar represents all events, in yellow. The last bar for each terrain type represents the $RD_{250}$ dataset (points along roads that are at least 250 meters from a known event) in blue. The x-axis is labeled with the terrain types, from 1 to 16, as proposed by Iwahashi and Pike in [22]. The magnitude of

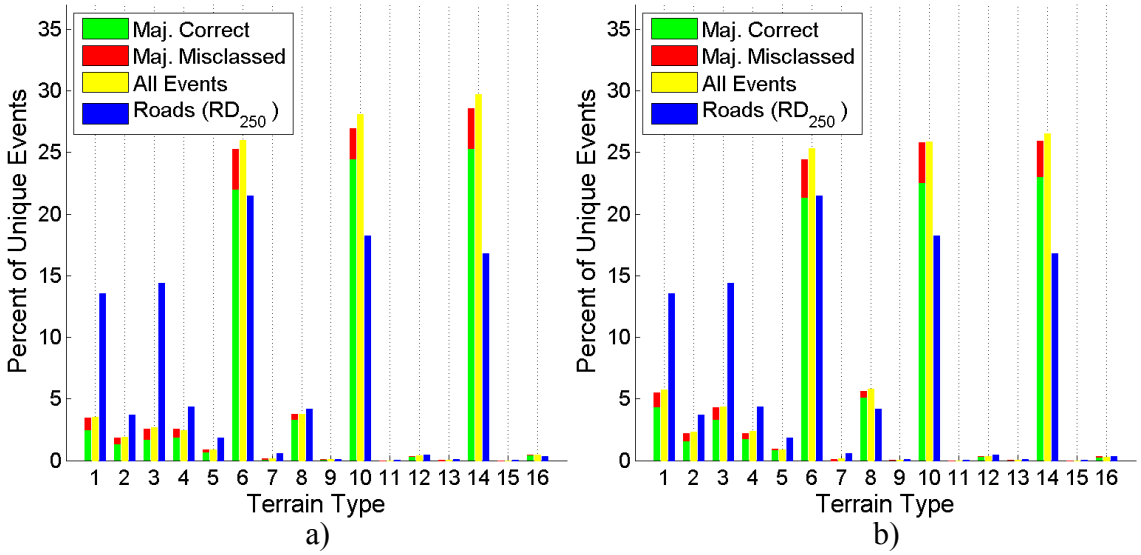



**Figure 63. Analysis of misclassified events, by terrain type; a) IED events; b) DF events.**

each bar is calculated as a percent of the total for that class. Thus, a blue bar with a magnitude of 8 indicates that 8% of all $RD_{250}$ points are of that particular terrain type.

In both figures, it is clear that the distribution of terrain types is not uniform in Afghanistan. Notably, terrain types 6, 10 and 14 —described in [22] as having lower gradient, coarse texture and high convexity— appear to dominate with most of the remaining points spread between types 1-4. This distribution is even more pronounced for events. In fact, the three most common terrain types include 56% of all $RD_{250}$ points, over 70% of all misclassified events (IED and DF), over 79% of all correctly classified events. Terrain types 1 and 3, especially prominent for roads, are described as high gradient, coarse texture and high convexity.

Figure 63.a shows the impact of terrain type on IED event classification. While misclassified events appear at several terrain types, they do not appear to be disproportionately represented in any particular terrain type. DF events show similar trends in Figure 62.b.

Figure 64 examines the impact of date or time of year on classification error using the commonly correct and common misclassified events as described above. These figures were constructed using counts collected over one-month windows. (The gaps are a reflection of an incomplete dataset. See Appendix A.2 for details.) Figure 64.a and 64.b examine the distribution of events and event error across the 19 months covered by the dataset, for IED and DF respectively. Each pair of bars represents the mean event counts of the individual classifiers (left, green/red) and the mean event counts of the ensemble classifiers (right, green/yellow) for a single month. For each bar, the green portion at the bottom is the count of correctly classified events. The top of the bar, in red or yellow, is the count of misclassified events. The numbers above each pair, labeled 'I' and 'E' are the percent of misclassified events for Individual classifiers and Ensemble classifiers respectively.

One interesting trend is the decrease in event classification error between May and August during both 2011 and 2012. While the reasons behind this trend are unclear, it may be that this is a reflection of the real-world learning that happens in the war zone in Afghanistan. Over the winter, both attackers and targets face periods of reduced mobility. This provides attackers time to develop new tactics which are then

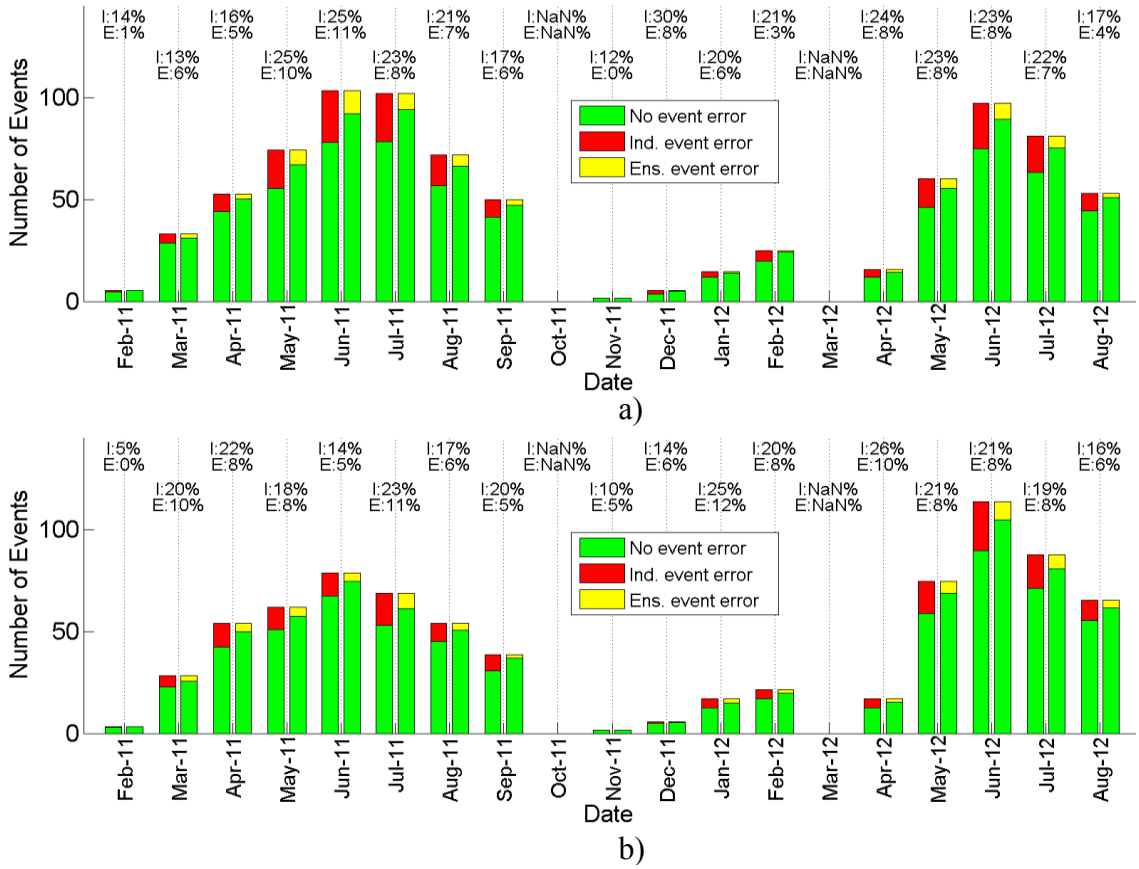



**Figure 64. Event classification errors by date; a) IED event classification errors by date and classifier type; b) DF event classification errors by date and classifier type.**

implemented in the spring as the temperatures go up. Targets are unable to develop counter-measures until the new tactics are seen. However, as the attacks ramp up and counter-measures are developed and disseminated, the number of successful attacks decreases. If this is true, then similar results might be expected from machine learning algorithms. Early in the season, there are fewer attacks and these may be a combination of old and new tactics. As the season progresses, more of the same tactics are seen and

the classifier ability to correctly classify them increases. This remains true until new tactics emerge.

Figure 65 examines the impact of time of day on classification error. The ratio of misclassified event to total events is explored in Figures 64.a and 64.b, for IED and DF events respectively. These figures were constructed using a one hour jumping window.

For IED events shown in Figure 65.a, there are general smaller errors seen late at night and early in the morning. While the reasons are unclear, this is possibly related to darkness. In the dark, it is more difficult for the attacker to see the target, so the range of sites available for an IED attack may be constrained.

Figure 65.b shows an opposite trend for direct fire attacks between the hours of 2300 and 0300, more easily seen in the numbers across the top of the figure. The event misclassification rate is much higher, almost doubled, for DF events that occur in this window. Although the numbers of events are relatively small, they still number in the hundreds, so a few bad entries in the dataset shouldn't have too large an effect. While the reason for this increase is unclear, it may be related to how the data is collected and logged. When a direct fire attack happens during daylight, it is normally possible to determine an approximate location for the shooter based on incoming trajectory, smoke and other factors related to shooting. However, at night it may be impossible to accurately estimate the actual location of the attack. This may lead to increased reporting errors at night.

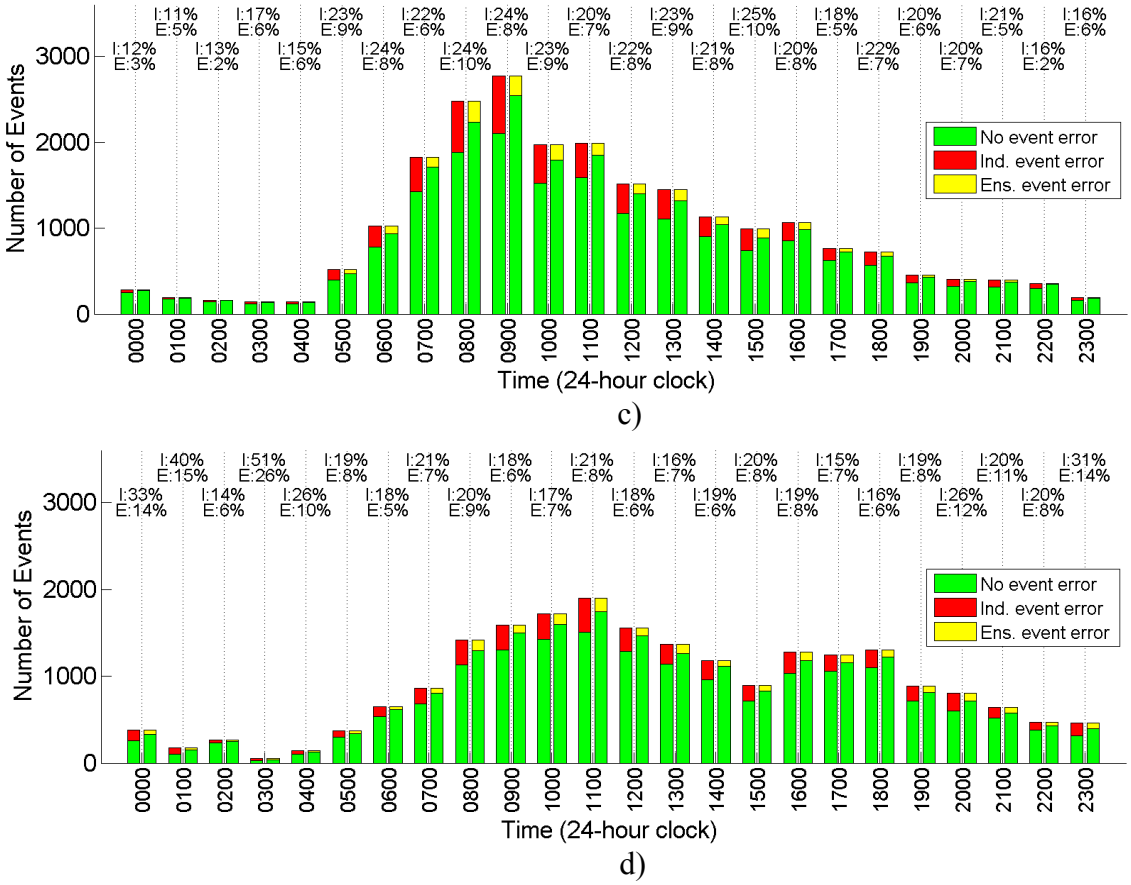
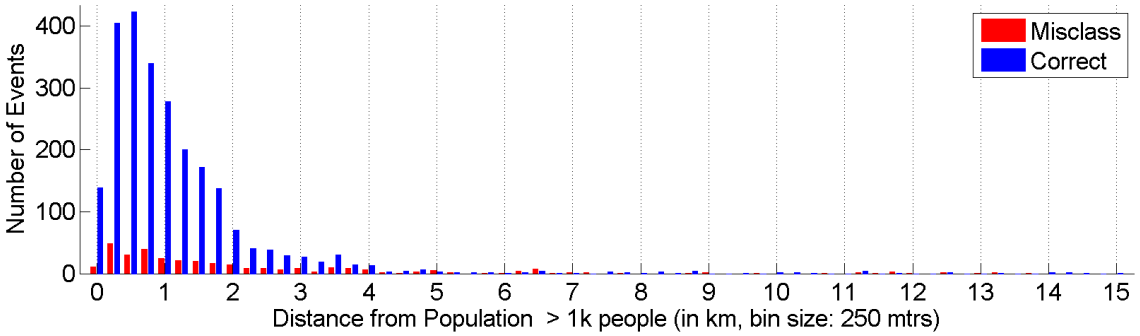
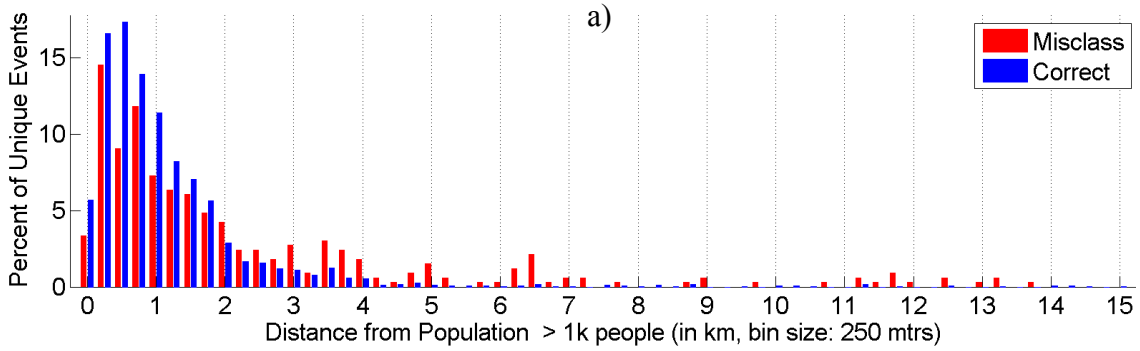



**Figure 65. Event classification errors by time of day; a) IED event classification errors by time of day and classifier type; b) DF event classification errors by time of day and classifier type.**

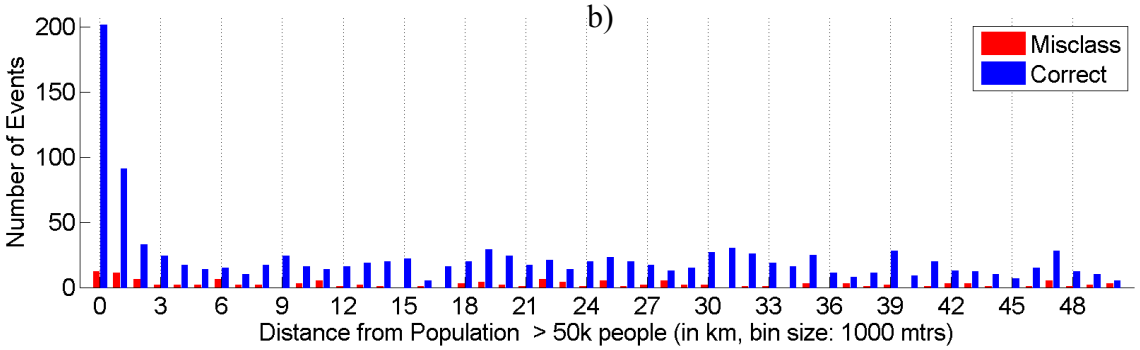
Distance to populated areas appears to have a little effect on correct classification of IED events, as shown in Figure 66. In the figure, the distance from populated areas to misclassified and correctly classified events is presented. Figure 66.a and 65.b examine the distance from populated areas with greater than 1000 people using a bin size, or increment, of 250 meters. The distribution of misclassified and correctly classified events is generally similar. One possibly significant difference is seen at distances greater than 6 km, where almost all (of the few) events are misclassified.



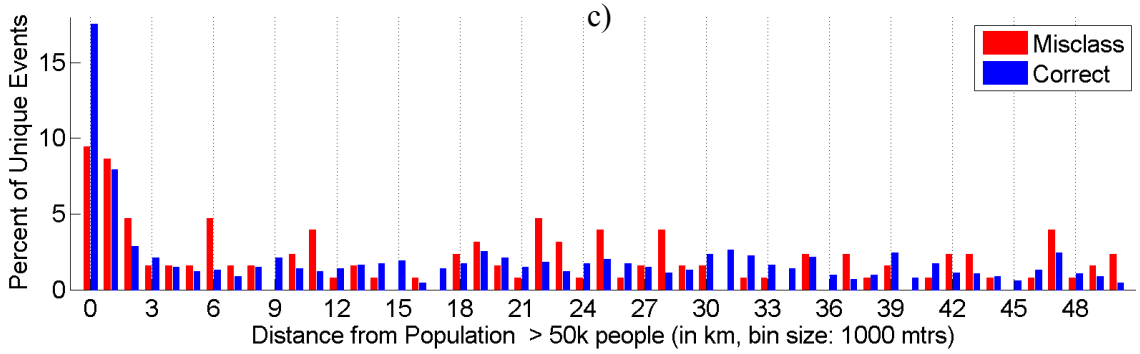

a)




b)




c)




d)

**Figure 66. IED event classification errors by distance from populated areas; a) error count by distance from populated areas with >1000 people; b) error percent by distance from populated areas with >1000 people; c) error count by distance from populated areas with >50,000 people; d) error percent by distance from populated areas with >50,000 people.**

Figures 66.c and 66.d offer a similar analysis for distances to populated areas with more than 50,000 people. As before, no obvious differences between misclassified and correctly classified events are obvious. A similar figure for DF events at the same distances can be found in Appendix D.

*Summary of Analysis of Event Classification Errors*

In the previous section, the performance of seven different classifiers was examined to determine if event classification error varied with regard to several different potentially strong influences. In each of the four cases —terrain type, time, date and distance from populated areas— no strong differences emerged from the comparison of misclassified and correctly classified events. In general, these two groups, misclassified and correctly classified events, tended to vary together over the domain of the influencing factor.

# CHAPTER VI

# CONCLUSION

In this work, we investigated and developed an underlying framework supporting predictive analysis of asymmetric conflict events. With the real world environment of modern day Afghanistan as a backdrop, we developed a model tailored to asymmetric conflict and populated it with conventional geomorphometric features as well as novel viewshed-based features that assess risk aversion and other human factors. Predictive analysis using these features produced classification error approaching 8%, even when trained with noisy data collected in unknown conditions. Due to the high cost of conflict event misclassification, a possible IED explosion or undetected ambush, the predictive analysis classifiers were assessed using a weighted cost metric that penalized event classification error.

MECH, the underlying asymmetric conflict model, was developed using two primary components: a model of the asymmetric warfare attack Emplacement site and a model of the Monitor and Control sites supporting the attack. A Halo constructed using the longest and shortest lines of sight geographically constrained the total area considered useful for Monitor and Control functions. This constraint informed the collection of features and defined the area to be analyzed. The resulting consolidated MECH model defined the attack site and its environs with a focus on geomorphometric and visibility-related features and human factors that support the execution of an asymmetric conflict event.

In order to populate the MECH model, conventional geomorphometric features were combined with a new set of features that attempt to capture the human factors of risk aversion and visibility of the event site. By identifying geographic locations within the Halo that simultaneously met two constraints, intervisibility with the target and adjacency to escape routes, these new features supported assessment of risk aversion and allowed the algorithmic quantification of locations with high utility for Monitor and Control activities. The resulting identification of optimal escape-adjacent points provided the basis for additional features assessing the visibility of approaches to an Emplacement site from highly useful and escape-adjacent Monitor and Control sites. A subset of these features, called route visibility in this research, have proven to be highly useful in the classification of conflict events.

Since humans involved in asymmetric conflict are unlikely to have the tools or training required to perform exhaustive terrain analysis supporting conflict event site selection, new features were proposed that use line-of-sight constraints and summarize key terrain descriptors a goal of measuring the 'feel' of a location. Features were developed that assess line of sight and openness of the terrain, describe the general shape and complexity of the viewshed, and summarize its texture or roughness. These features, constructed by analyzing a limited number of radials centered on the Emplacement site, provide computationally inexpensive insight into the general utility of terrain for a conflict event.

Given the MECH model populated with potentially relevant features, learning algorithms were trained with a goal of correctly classifying points along roads into two

classes: Events and Non-Events. The topics of data normalization and feature selection were explored and choices were made tailored to this specific data set. Machine learning algorithms including Support Vector Machines with various kernels, k-Nearest Neighbors, and both Linear and Quadratic Discriminant Analysis were employed under a variety of geographic and temporal constraints, both fixed and dynamic, and under various subset selection rules, including balanced and unbalanced classes. Feature selection by experts was examined and found to be inferior to machine learning approaches. Learning algorithm refinement was performed using dynamic parameter and hyperparameter estimation for both kNN and SVM. Further improvement was discovered through the use of ensembles.

In the end, the best performing unconstrained classifier used SVM with dynamic parameter estimation to train on a stack consisting of ten mixed base-classifiers configured as a stack. In four of the individual classifiers that made up the stack, features were divided using the MECH model as a guide and normalized using event class moments determined dynamically from the training set. The order of $k$ was dynamically selected for kNN using the training set. Both the RBF shape parameter $\sigma$ and the box constraint, $C$, are dynamically estimated for SVM from the training set using a two-stage grid search. This classification construct produced overall classification error of less than 10%. More importantly, event misclassification was reduced to less than 5% of all events in the sample for both IED and DF.

Always keeping in mind that the end user of this research is likely to be a soldier on a modern battlefield, light-weight classification solutions were also explored. Various

solutions, designed to be used in resource-constrained environments, were examined, with a primary focus on the use of single classifiers and computationally cheap features. The best results were obtained using SVM on a reduced, fixed set of 17 features. With an overall classification error around 24% and event misclassification of 10.2% of all events in the sample, the resource-constrained version provided surprisingly good results.

Experimental results derived from real-world asymmetric conflict data in Afghanistan show that the MECH model and the features developed in this research are useful in the predictive analysis of asymmetric conflict events. The proposed modeling technique and classification solutions provide a concrete foundation for the development of practical tools and applications supporting situational awareness in asymmetric conflict environments.

In the course of this research a variety of constraining choices were made. Further exploration of these choices may result in better features or more accurate classification. One of the most interesting choices dealt with time.

The importance of time emerges is visible in several ways. First, a training set duration of 120 days was selected. This was paired with a test set duration of 60 days. Thus, the scale of both the training and test sets were fixed. In the case of the training set, this length of time is closely correlated with seasonal weather. Future work might consider alternate and potentially more dynamic ways of selecting a temporal constraint. Also, seasonal trends are probably important. In fact, it is likely that cold weather and hot weather drive key decisions regarding tactics. The impact of seasonal weather may

also play a key factor. Thus, correlating tactical patterns across seasons might be a useful way to select more homogeneous training sets. Finally, time of day is not currently a consideration for viewshed calculations. However, surface irradiance, heat convection, wind, and localized weather patterns may all depend on or correlate with time of day or position of the sun. Anisotropic surface considerations, particularly with respect to irradiance and shadows, are not currently considered. It is likely that real viewsheds are significantly smaller, particularly at longer ranges, when these factors are considered.

Geomorphometric choices are another area where potential improvements exist. These improvements are probably related to algorithmic choices, scale, and error management. In the area of geomorphometry, this research chose to implement algorithms commonly mentioned in academic papers or discussed in textbooks. For example, surface roughness is described as a simple standard deviation from mean elevation. However, for most of the features captured here, other algorithms exist that may quantify the land surface in ways that more accurately capture its structure or utility. Scale is also important. For many features, the scale was arbitrarily fixed to area within some specified radius of a fixed point. However, it is likely that some features will benefit from smaller, larger or dynamic scales. The resolution of the elevation data was an important constraint in this area. Built on a 30-meter grid, any feature smaller than one pixel was probably invisible or diluted by the DEM spacing. Finally, DEM error is probably an important contributor to the noise found in some of the features. Error correction strategies should probably be an important part of any feature that uses elevation data.

# REFERENCES


[1] United States. Dept. of the Army, *Military Operations in Low Intensity Conflict.* Washington, D.C.: Headquarters, Depts. of the Army and the Air Force, 1990.

[2] "IED Deaths in Operation Enduring Freedom | Afghanistan." [Online]. Available: http://icasualties.org/oef/. [Accessed: 02-June-2014].

[3] W. L. Perry and J. Gordon, "Analytic Support to Intelligence in Counterinsurgencies," Santa Monica, CA : RAND National Defense Research Institute, 2008.

[4] M. Townsley, S. D. Johnson, and J. H. Ratcliffe, "Space Time Dynamics of Insurgent Activity in Iraq," *Security Journal*, vol. 21, no. 3, pp. 139–146, Jul. 2008.

[5] A. Gutfraind, "Mathematical Terrorism: Quantitative Modeling of Sub-state Conflicts," Thesis (Ph.D.), Cornell University, Feb. 2010..

[6] N. Memon, J. D. Farley, D. L. Hicks, and T. Rosenorn, Eds., *Mathematical Methods in Counterterrorism*, Wien: Springer-Verlag, 2009.

[7] J. A. Lynn, "Patterns of Insurgency and Counterinsurgency," *Military Review*, vol. 85, no. 4, p. 22-27, Jul. 2005.

[8] "Defeating The Improvised Explosive Device (IED) and Other Asymmetric Threats Today's Efforts and Tomorrow's Requirements: Hearing before the Oversight and Investigations Subcommittee of the Committee on Armed Services, House of Representatives, One Hundred Tenth Congress, Second Session, Hearing held

September 16, 2008.," United States Congress House Committee on Armed Services Oversight and Investigations Subcommittee, Washington D.C., 2009.

[9] A. D. Barker, "Improvised Explosive Devices in Southern Afghanistan and Western Pakistan, 2002–2009," *Studies in Conflict & Terrorism*, vol. 34, no. 8, pp. 600–620, 2011.

[10] S. Magnuson, "Adaptive Foe Thwarts Counter-IED Efforts: Pentagon's Efforts to Neutralize Roadside Bombs Are Stymied by a Nimble Enemy," *National Defense*, vol. 90, pp. 22–23, 2006.

[11] M. A. Hanson, "The Economics of Roadside Bombs," Department of Economics, College of William and Mary, Working Paper 68, 2007.

[12] C. Wilson, "Improvised Explosive Devices (IEDs) in Iraq and Afghanistan: Effects and Countermeasures," Washington, D.C.: Congressional Research Service, Library of Congress, 2007. [Online]. Available: http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA475029. [Accessed: 02-June-2014].

[13] *IED Trigger Recognition Guide*, 3rd ed. Washington D.C.: National Counterterrorism Center, 2006.

[14] C. Partelow, "Using GIS to Depict Resource Risk from Probable Cannabis Cultivation Sites," Thesis (M.A.), San Jose State University, 2008.

[15] M. Mallery, "Marijuana National Forest: Encroachment on California Public Lands for Cannabis Cultivation," *Berkeley Undergraduate Journal*, vol. 23, no. 2, pp. 1–50, 2011.

[16] J. M. Caplan and L. W. Kennedy, *Risk Terrain Modeling Manual : Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis*. Newark, NJ: Rutgers Center on Public Security, 2010.

[17] *Ranger Handbook*. Fort Benning, GA: United States Army Infantry School, Ranger Department, 2013.

[18] *Counterinsurgency Field Manual (Army FM 3-24)*. Washington, D.C.: Headquarters, Dept. of the Army, 2007.

[19] M. Janlov, T. Salonen, H. Seppanen, and K. Virrantaus, "Developing Military Situation Picture by Spatial Analysis and Visualization," presented at the ScanGIS 2005: The 10th Scandinavian Research Conference on Geographical Information Science, Stockholm, Sweden, 2005.

[20] R. J. Pike, I. S. Evans, and T. Hengl, "Geomorphometry: A Brief Guide," in *Developments in Soil Science*, vol. 33, Tomislav Hengl and Hannes I. Reuter, Ed. Elsevier, pp. 3–30, 2009.

[21] R. J. Pike, "The Geometric Signature: Quantifying Landslide-terrain Types from Digital Elevation Models," *Mathematical Geology*, vol. 20, no. 5, pp. 491–511, 1988.

[22] J. Iwahashi and R. J. Pike, "Automated Classifications of Topography from DEMs by an Unsupervised Nested-means Algorithm and a Three-part Geometric Signature," *Geomorphology.*, vol. 86, no. 3, pp. 409-440, 2007.

[23] T. F. Stepinski and C. Bagaria, “Segmentation-Based Unsupervised Terrain Classification for Generation of Physiographic Maps,” *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 733–737, Oct. 2009.

[24] L. Huaxing, “Modelling Terrain Complexity,” in *Advances in Digital Terrain Analysis*, Q. Zhou, B. Lees, and G. Tang, Eds. Springer Berlin Heidelberg, pp. 159–176, 2008.

[25] Y. H. Chou, P. S. Liu, and R. J. Dezzani, “Terrain Complexity and Reduction of Topographic Data,” *Journal of Geographical Systems*, vol. 1, no. 2, pp. 179–198, 1999.

[26] P. T. Giles, “Geomorphological Signatures: Classification of Aggregated Slope Unit Objects from Digital Elevation and Remote Sensing Data,” *Earth Surface Processes and Landforms*, vol. 23, no. 7, pp. 581–594, 1998.

[27] J. M. Azanon, J. Delgado, and A. Gomez, “Morphological Terrain Classification and Analysis Using Geostatistical Techniques,” *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 35, pp. 62–67, 2004.

[28] J. E. Vogelmann, T. Sohl, and S. M. Howard, “Regional Characterization of Land Cover Using Multiple Sources of Data,” *Photogrammetric Engineering and Remote Sensing*, vol. 64, no. 1, p. 45-48, 1998.

[29] R. Richbourg and W. K. Olson, “A Hybrid Expert System that Combines Technologies to Address the Problem of Military Terrain Analysis,” *Expert Systems with Applications*, vol. 11, no. 2, p. 207-225, 1996.

[30] G. Nagy, “Terrain Visibility,” *Computers & Graphics*, vol. 18, no. 6, pp. 763–773, Nov. 1994.

[31] L. D. Floriani and P. Magillo, “Intervisibility on Terrains,” in *Geographic Information Systems: Principles, Techniques, Management and Applications*, 2nd ed., vol. 2, 2 vols., John Wiley & Sons, pp. 543–556, 1999.

[32] L. D. Floriani and P. Magillo, “Algorithms for Visibility Computation on Terrains: a Survey,” *Environment and Planning B: Planning and Design*, vol. 30, no. 5, pp. 709 – 728, 2003.

[33] “Line of Sight Compendium,” U.S. Army Corps of Engineers, 2004.

[34] W. R. Franklin, “Siting Observers on Terrain,” in *Advances in Spatial Data Handling*, Ottawa, CA, 109-120, 2002.

[35] P. N. Stiles, “Terrain intervisibility -Believe it or not?,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 16, no. 8, pp. 25–30, Aug. 2001.

[36] E. Moet, M. Van Kreveld, and R. Van Oostrum, “Region Intervisibility in Terrains,” *International Journal of Computational Geometry & Applications*, vol. 17, no. 04, pp. 331–347, 2007.

[37] P. Guth, *Studies in Military Geography and Geology*. Kluwer, Netherlands: Geological Society of America Reviews in Engineering Geology, 2004.

[38] S. M. Raehtz, “Accounting for Uncertainty in Viewshed Analysis of IED Ambush Sites in Afghanistan,” Thesis (M.S.), Michigan State University, 2011.

[39] P. Guth, “Shortcuts in the Line-of-sight and Viewshed Algorithms with Gridded Geographic DEMs: What’s Permissible and What’s Illegitimate,” in *American*

*Society of Photogrammetry and Remote Sensing 2005 Conference Proceedings*, Baltimore, MD, 2005.

[40] D. Izraelevitz, "A Fast Algorithm for Approximate Viewshed Computation," *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 7, pp. 767–774, Jul. 2003.

[41] J. Washtell, S. Carver, and K. Arrell, "A Viewshed Based Classification of Landscapes Using Geomorphometrics," in *Geomorphometry 2009 Conference Proceedings*, Zurich, Switzerland, pp. 44–49, 2009.

[42] J. K. Rød and D. van der Meer, "Visibility and Dominance Analysis: Assessing a High-rise Building Project in Trondheim," *Environment and Planning B: Planning and Design*, vol. 36, no. 4, pp. 698 – 710, 2009.

[43] T. Hengl and H. I. Reuter, *Geomorphometry: Concepts, Software, Applications*. Oxford: Elsevier, 2009.

[44] R. Yokoyama, M. Shirasawa, and R. Pike, "Visualizing Topography by Openness: A New Application of Image Processing to Digital Elevation Models," *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 3, pp. 251–266, 2002.

[45] P. Guth, A. Nantet, and A. Schaeffer, "Using Upward Openness for Viewshed Prediction," in *Proceedings of 75th Annual Conference of the American Society for Photogrammetry and Remote Sensing*, Baltimore, MD, 2009.

[46] M. L. Benedikt, "To Take Hold of Space: Isovists and Isovist Fields," *Environment and Planning B: Planning and Design*, vol. 6, no. 1, pp. 47 – 65, 1979.

[47] A. Turner, “From Isovists to Visibility Graphs: A Methodology for the Analysis of Architectural Space,” *Journal of Planning Literature*, vol. 16, no. 1, pp. 80–163, 2001.

[48] D. Wheatley, “Cumulative Viewshed Analysis: a GIS-based Method for Investigating Intervisibility, and its Archaeological Application,” in *Archaeology and Geographical Information Systems: A European Perspective*, CRC Press, pp. 171-185, 1995.

[49] M. Lake, P. Woodman, and S. Mithen, “Tailoring GIS software for Archaeological Applications: an Example Concerning Viewshed Analysis,” *Journal of Archaeological Science*, vol. 25, no. 1, pp. 27–38, 1998.

[50] K. Curtin and N. Perry, “Linear Referencing for Network Analysis of IED Events,” presented at the Armed Forces Communications and Electronics Association (AFCEA) Symposium, Lansdowne, VA, 2009.

[51] J. E. Alt, G. King, and C. S. Signorino, “Aggregation among Binary, Count, and Duration Models: Estimating the Same Quantities from Different Levels of Data,” *Political Analysis*, vol. 9, no. 1, pp. 21–44, Jan. 2001.

[52] R. C. Holte, “Very Simple Classification Rules Perform Well on Most Commonly Used Datasets,” *Machine Learning*, vol. 11, no. 1, p. 63-90, 1993.

[53] O. Stone, *Platoon*. Hemdale Film Corporation, 1986.

[54] M. Townsend, “Point Man in Afghanistan: a Soldier’s View,” *The Guardian*, London, UK, 01-Apr-2012.

[55] M. Gladwell, *Blink: the Power of Thinking Without Thinking*. New York: Little, Brown and Co., 2005.

[56] G. Zacharias, J. MacMillan, and S. Hemel, Eds., *Behavioral Modeling and Simulation: From Individuals to Societies*. Washington D.C.: National Academies Press, 2008.

[57] S. J. Deitchman, "A Lanchester Model of Guerrilla Warfare," *Operations Research*, vol. 10, no. 6, pp. 818–827, Nov. 1962.

[58] M. Schaffer, "Lanchester Models of Guerrilla Engagements," *Operations Research*, vol. 16, no. 3, pp. 457–488, 1968.

[59] M. Kress and R. Szechtman, "Why Defeating Insurgencies is Hard: The Effect of Intelligence in Counterinsurgency Operations — A Best-case Scenario," *Operations Research*, vol. 57, no. 3, pp. 578–585, May 2009.

[60] Z. A. Henscheid, M. T. Koehler, S. K. Mulutzie, B. F. Tivnan, and J. G. Turnley, "COIN 2.0 Formulation," Mitre Corporation, Bedford, MA, 2010.

[61] A. Bornstein, T. Damarla, J. Lavery, F. Morelli, and E. Schmeisser, "Remote Detection of Covert Tactical Adversarial Intent of Individuals in Asymmetric Operations," Army Research Laboratory, Computational and Information Sciences Directorate, Aberdeen Proving Grounds, MD, ARL-SR-197, 2010.

[62] P. Shakarian, M. Nagel, B. Schuetzle, and V. Subrahmanian, "Abductive Inference for Combat: Using SCARE-S2 to Find High-value Targets in Afghanistan," in *Proceedings of the 23rd Innovative Applications of Artificial Intelligence Conference*, San Francisco, CA, 2011.

[63] N. Johnson, S. Carran, J. Botner, K. Fontaine, N. Laxague, P. Nuetzel, J. Turnley, and B. Tivnan, "Pattern in Escalations in Insurgent and Terrorist Activity," *Science*, vol. 333, no. 6038, pp. 81–84, Jul. 2011.

[64] N. F. Johnson, "Escalation, Timing and Severity of Insurgent and Terrorist Events: Toward a Unified Theory of Future Threats," 09-Sep-2011. [Online]. Available: http://ssrn.com/abstract=2003876. [Accessed: 10-May-2014].

[65] H. Okuda, S. Hayakawa, T. Suzuki, and N. Tsuchida, "Modeling of Human Behavior in Man-machine Cooperative System Based on Hybrid System Framework," in *Proceeding of the 2007 IEEE International Conference on Robotics and Automation*, Roma, Italy, pp. 2534–2539, 2007.

[66] A. Pentland and A. Liu, "Modeling and Prediction of Human Behavior," *Neural Computation*, vol. 11, no. 1, pp. 229–242, Jan. 1999.

[67] C. L. Barrett, S. Eubank, V. A. Kumar, and M. V. Marathe, "Understanding Large Scale Social and Infrastructure Networks: A Simulation-based Approach," *Society for Industrial and Applied Mathematics News*, vol. 37, no. 4, pp. 1–5, 2004.

[68] C. Barrett, S. Eubank, and M. Marathe, "Modeling and Simulation of Large Biological, Information and Socio-technical Systems: an Interaction-based Approach," in *Interactive Computation*, Springer, pp. 353–392, 2006.

[69] A. Wilson, G. Wilson, and D. H. Olwell, Eds., *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, 1st ed. New York: Springer, 2006.

[70] T. Sandler and K. Siqueira, “Games and Terrorism Recent Developments,” *Simulation Gaming*, vol. 40, no. 2, pp. 164–192, Apr. 2009.

[71] J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, and N. F. Johnson, “Common Ecology Quantifies Human Insurgency,” *Nature*, vol. 462, no. 7275, pp. 911–914, Dec. 2009.

[72] A. Washburn and L. P. Lee Ewing, “Allocation of Clearance Assets in IED Warfare,” *Naval Research Logistics*, vol. 58, no. 3, pp. 180–187, 2011.

[73] A. Dixon, Z. Zhao, J. C. Bohorquez, R. Denney, and N. Johnson, “Statistical Physics and Modern Human Warfare,” in *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, G. Naldi, L. Pareschi, and G. Toscani, Eds. Birkhäuser Boston, pp. 365–396, 2010.

[74] C. Fâbian and Z. Szo‘ke, “Solving Two-stage Stochastic Programming Problems with Level Decomposition,” *Computational Management Science*, vol. 4, no. 4, pp. 313–353, 2007.

[75] G. Jakobson, J. Buford, and L. Lewis, “Situation Management: Basic Concepts and Approaches,” in *Information Fusion and Geographic Information Systems*, Springer, pp. 18–33, 2007.

[76] G. Jakobson, J. Buford, and L. Lewis, “Towards an Architecture for Reasoning about Complex Event-based Dynamic Situations,” in *Proceedings of the 26th International Conference on Software Engineering*, Edinburgh, UK, pp. 62–67, 2004.

[77] E. Nikolova, “Approximation Algorithms for Offline Risk-averse Combinatorial Optimization,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Berlin, Germany: Springer Berlin Heidelberg, pp. 338–351, 2010.

[78] C. Swamy, “Risk-averse Stochastic Optimization: Probabilistically-constrained Models and Algorithms for Black-box Distributions,” in *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, CA, pp. 1627–1646, 2011.

[79] V. Martinez, G. Simari, A. Sliva, and V. Subrahmanian, “CONVEX: Similarity-based Algorithms for Forecasting Group Behavior,” *Intelligent Systems, IEEE*, vol. 23, no. 4, pp. 51–57, 2008.

[80] V. Subrahmanian, M. Albanese, M. Martinez, D. Nau, and D. Reforgiato, “CARA: A Cultural-reasoning Architecture,” *Intelligent Systems, IEEE*, vol. 22, no. 2, pp. 12–16, 2007.

[81] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag, 2001.

[82] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from Data: A Short Course*, 1st ed. Albany, NY: AMLBook (EMSQRD LLC), 2012.

[83] A. K. Jain, R. P. Duin, and J. Mao, “Statistical Pattern Recognition: a Review,” *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[84] R. T. Ng and J. Han, "CLARANS: a Method for Clustering Objects for Spatial Data Mining," *IEEE Transactions On Knowledge And Data Engineering*, vol. 14, no. 5, pp. 1003–1016, 2002.

[85] R. Ng and J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," in *Proceedings of the 20th International Conference on Very Large Data Bases*, Santiago de Chile, Chile, pp. 144-155, 1994.

[86] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi, "Exploratory Spatio-temporal Data Mining and Visualization," *Journal of Visual Languages and Computing*, vol. 18, no. 3, pp. 255–279, 2007.

[87] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal Clustering: a Survey," in *Data Mining and Knowledge Discovery Handbook*, Springer, pp. 855–874, 2010.

[88] A. Globerson and S. Roweis, "Metric Learning by Collapsing Classes," in *Advances in Neural Information Processing Systems 18*, Cambridge, MA: MIT Press, pp. 451–458, 2006.

[89] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood Components Analysis," in *Advances in Neural Information Processing Systems 17*, MIT Press, pp. 513–520, 2004.

[90] Z. Yang and J. Laaksonen, "Regularized Neighborhood Component Analysis," in *Proceedings of the 15th Scandinavian Conference on Image Analysis*, Aalborg, Denmark, pp. 253–262, 2007.

[91] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. New York: Springer, 2006.

[92] C. Cortes and V. Vapnik, "Support-vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[93] G. Mountrakis, J. Im, and C. Ogole, "Support Vector Machines in Remote Sensing: a Review," *International Society for Photogrammetry and Remote Sensing Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[94] C. Yeh, W. Su, and S. Lee, "Employing Multiple-kernel Support Vector Machines for Counterfeit Banknote Recognition," *Applied Soft Computing*, vol. 11, no. 1, pp. 1439–1447, Jan. 2011.

[95] T. Marwala and M. Lagazio, "Support Vector Machines for Modeling Interstate Conflict," in *Militarized Conflict Modeling Using Computational Intelligence*, Springer London, pp. 89–105, 2011.

[96] C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[97] C. Chang and C. Lin, "LIBSVM: a Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, p. 27, 2011.

[98] C. Hsu, C. Chang, and C. Lin, "A Practical Guide to Support Vector Classification," Department of Computer Science, National Taiwan University, 2003. [Online]. Available: https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf. [Accessed: 02-June-2014].

[99] X. Wang, X. Chen, and M. Li, “Classification of High Spatial Resolution Remote Sensing Image using SVM and Local Spatial Statistics Getis-Ord Gi,” in *Proceedings of SPIE Volume 8006*, 2011.

[100] G. Zhu and D. G. Blumberg, “Classification Using ASTER data and SVM Algorithms;: The Case Study of Beer Sheva, Israel,” *Remote Sensing of Environment*, vol. 80, no. 2, pp. 233–240, May 2002.

[101] P. Goncalves, H. Carrão, A. Pinheiro, and M. Caetano, “Land Cover Classification with Support Vector Machine Applied to MODIS Imagery,” in *Proceedings of the 25th European Association of Remote Sensing Laboratories Symposium*, Porto, Portugal, pp. 6–11, 2005.

[102] S. S. Durduran, “A Decision Making System to Automatically Recognize Traffic Accidents on the Basis of a GIS platform,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 7729–7736, Dec. 2010.

[103] M. Kanevski, M. Maignan, E. Mayoraz, and N. Gilardi, “Environmental and Pollution Spatial Data Classification with Support Vector Machines and Geostatistics,” in *Proceedings of ACAI-99 Workshop on Machine Learning and Intelligent Agents*, Chania, Crete, pp. 43–51, 1999.

[104] C. H. Li, C. T. Lin, B. C. Kuo, and H. S. Chu, “An Automatic Method for Selecting the Parameter of the RBF Kernel Function to Support Vector Machines,” in *Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium*, Honolulu, HI, pp. 836–839, 2010.

[105] M. Mongillo, "Choosing Basis Functions and Shape Parameters for Radial Basis Function Methods," *SIAM Undergraduate Research Online,* vol. 4, 190-209, 2011.

[106] Z. Liu, M. J. Zuo, and H. Xu, "Parameter Selection for Gaussian Radial Basis Function in Support Vector Machine Classification," in *Proceedings of the 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, Chengdu, China, pp. 576–581, 2012.

[107] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[108] J. He, S. Kumar, and S. Chang, "On the Difficulty of Nearest Neighbor Search," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, Edinburgh, Scotland, pp. 1127–1134, 2012.

[109] T. K. Ho, "Nearest Neighbors in Random Subspaces," in *Advances in Pattern Recognition*, vol. 1451, Sydney, Australia: Springer, 1998, pp. 640–648.

[110] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[111] K. M. Ting and I. H. Witten, "Issues in Stacked Generalization," *Journal of Artificial Intelligence Research*, vol. 10, pp. 271–289, 1999.

[112] D. W. Fan, P. K. Chan, and S. J. Stolfo, "A Comparative Evaluation of Combiner and Stacked Generalization," in *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Portland, OR, pp. 40–46, 1996.

[113] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[114] J. N. Arvesen and G. P. McCabe, “Subset Selection Problems for Variances with Applications to Regression Analysis,” *Journal of the American Statistical Association*, vol. 70, no. 349, pp. 166–170, Mar. 1975.

[115] S. Gupta and D. Huang, “On Selecting an Optimal Subset of Regression Variables,” Defense Technical Information Center, Purdue University Department of Statistics, ADA044889, 1977.

[116] A. J. Miller, “Selection of Subsets of Regression Variables,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 147, no. 3, pp. 389-425, 1984.

[117] P. L. Flom and D. L. Cassell, “Stopping Stepwise: Why Stepwise and Similar Selection Methods are Bad, and What You Should Use,” presented at the NorthEast SAS Users Group Inc 20th Annual Conference:, Baltimore, Maryland, 2007.

[118] R. van den Berg, H. Hoefsloot, J. Westerhuis, A. Smilde, and M. van der Werf, “Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data,” *BioMed Central Genomics*, vol. 7, no. 1, 2006.

[119] A. Jain, K. Nandakumar, and A. Ross, “Score Normalization in Multimodal Biometric Systems,” *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[120] J. Karvanen, “The Statistical Basis of Laboratory Data Normalization,” *Drug Information Journal*, vol. 37, no. 1, pp. 101–107, Jan. 2003.

[121] E. Ogasawara, L. C. Martinez, D. de Oliveira, G. Zimbrão, G. L. Pappa, and M. Mattoso, “Adaptive Normalization: A Novel Data Normalization Approach for

Non-stationary Time Series," in *Proceedings of the 2010 International Joint Conference on Neural Networks*, pp. 1–8, 2010.

[122] G. King and L. Zeng, "Logistic Regression in Rare Events Data," *Political Analysis*, vol. 9, no. 2, pp. 137–163, Jan. 2001.

[123] Y. Huang, J. Pei, and H. Xiong, "Mining Co-Location Patterns with Rare Events from Spatial Data Sets," *GeoInformatica*, vol. 10, no. 3, pp. 239–260, 2006.

[124] *Army Field Manual: Tactics (FM 3-90)*. Washington D.C.: Headquarters, US Army Training and Doctrine Command, 2001.

[125] A. Apolloni, K. Channakeshava, L. Durbeck, M. Khan, C. Kuhlman, B. Lewis, and S. Swarup, "A Study of Information Diffusion Over a Realistic Social Network Model," in *Proceedings of the 12th International Conference on Computational Science and Engineering*, vol. 4, Vancouver, Canada, pp. 675–682, 2009.

[126] J. H. Norwitz, *Armed Groups: Studies In National Security, Counterterrorism, And Counterinsurgency*. Newport, RI: Naval War College Press, 2008.

[127] NASA Land Processes Distributed Active Archive Center, "ASTER GDEM Data Download Site." [Online]. Available: https://lpdaac.usgs.gov/get_data. [Accessed: 12-Nov-2013].

[128] R. A. MacMillan and P. A. Shary, "Landforms and Landform Elements in Geomorphometry," in *Geomorphometry — Concepts, Software, Applications*, vol. 33, Tomislav Hengl and Hannes I. Reuter, Ed. Elsevier, pp. 227–254, 2009.

[129] C. J. Moran and E. N. Bui, "Spatial Data Mining for Enhanced Soil Map Modelling," *International Journal of Geographical Information Science*, vol. 16, no. 6, pp. 533–549, 2002.

[130] P. Guth, J. Ehlen, and R. Harmon, "Quantifying Terrain Fabric in Digital Elevation Models," in *The Environmental Legacy of Military Operations*, vol. 14, Geological Society of America, 2001.

[131] M. V. Kreveld, "Variations on Sweep Algorithms: Efficient Computation of Extended Viewsheds and Class Intervals," in *Proceedings of the 7th International Symposium on Spatial Data Handling*, Delft, NL, pp. 13–15, 1996.

[132] Y. Shen, L. Lin, M. Yang, and G. Yurong, "Viewshed Computation Based on LOS Scanning," in *2008 International Conference on Computer Science and Software Engineering,* , vol. 2, pp. 984 –987, 2008.

[133] B. K. Horn, "Hill Shading and the Reflectance Map," *Proceedings of the IEEE*, vol. 69, no. 1, pp. 14–47, 1981.

[134] L. Zelnik-Manor and P. Perona, "Self-tuning Spectral Clustering," in *Advances in Neural Information Processing Systems 17*, pp. 1601–1608, 2004.

[135] M. Lagrange, L. G. Martins, and G. Tzanetakis, "Cluster-aware Normalization for Enhancing Audio Similarity," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, pp. 1969–1972, 2012.

[136] R. R. Hocking, "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, vol. 32, no. 1, pp. 1-49, Mar. 1976.

[137] *Matlab 2012b*. Natick, Massachusetts, USA: The Mathworks, Inc., 2012.

[138] K. Kira and L. A. Rendell, "A Practical Approach to Feature Selection," in *Proceedings of the Ninth International Workshop on Machine Learning*, Aberdeen, Scotland, UK, pp. 249–256, 1992.

[139] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral Region Classification Using a Three-Dimensional Gabor Filterbank," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 9, pp. 3457–3464, 2010.

[140] J. Shlens, "A Tutorial on Principal Component Analysis," *Systems Neurobiology Laboratory, University of California at San Diego,* 82, 2005. [Online]. Available: http://www.brainmapping.org/NITP/PNA/Readings/pca.pdf. [Accessed: 02-June-2014].

[141] C. J. Burges, "Dimension Reduction: A Guided Tour," *Machine Learning*, vol. 2, no. 4, pp. 275–365, 2009.

[142] I. Jolliffe, "Principal Component Analysis," in *Encyclopedia of Statistics in Behavioral Science*, vol. 3, John Wiley & Sons, Ltd, pp. 1580–1584, 2005.

[143] A. K. Ghosh, "On Nearest Neighbor Classification Using Adaptive Choice of k," *Journal of Computational and Graphical Statistics*, vol. 16, no. 2, pp. 482–502, 2007.

[144] G. G. Enas and S. C. Choi, "Choice of the Smoothing Parameter and Efficiency of k-nearest Neighbor Classification," *Computers & Mathematics with Applications*, vol. 12, no. 2A, pp. 235–244, 1986.

[145] P. Hall, B. Park, and R. Samworth, "Choice of Neighbor Order in Nearest Neighbor Classification," *Annals of Statistics*, vol. 36, no. 5, pp. 2135–2152, 2008.

[146] D. J. Hand and V. Vinciotti, "Choosing k for Two-class Nearest Neighbour Classifiers with Unbalanced Classes," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1555–1562, 2003.

[147] H. Dongyuan and C. Xiaoyun, "Tuning SVM Hyperparameters in the Primal," in *Proceedings of the Second International Conference on Computational Intelligence and Natural Computing Proceedings*, vol. 1, Wuhan, China, pp. 201–204, 2010.

[148] H. Han, W. Wang, and B. Mao, "Borderline-SMOTE: a New Over-sampling Method in Imbalanced Data Sets Learning," in *Advances in Intelligent Computing*, Springer Berlin Heidelberg, pp. 878–887, 2005.

[149] Q. Wei and R. L. Dunbrack Jr, "The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013.

[150] L. Rokach, "Ensemble-based Classifiers," *Artificial Intelligence Review*, vol. 33, no. 1–2, pp. 1–39, Feb. 2010.

[151] L. Rokach, "Genetic Algorithm-based Feature Set Partitioning for Classification Problems," *Pattern Recognition*, vol. 41, no. 5, pp. 1676–1700, May 2008.

[152] L. Kuncheva and C. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181–207, May 2003.

# APPENDIX A

The following sections describe the data used for this project.

## Appendix A.1 Global Digital Elevation Model

Elevation maps were obtained from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 [127]. Dated October 2011, these maps offer digital elevations with a horizontal resolution of approximately 30 meters (1/3 arc second). The data uses the WGS84 geoid and is stored in GeoTIFF format in 1° x 1° tiles. The ASTER L1B data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/get_data) .

The absolute vertical error for this product is estimated to be ±17 meters. However, the relative error (between adjacent pixels) is much smaller. At the distances used for MECH analysis, on the order of 3 km, it is unlikely that this error is problematic.

A more serious problem is the resolution. Each pixel in these elevation maps covers a ~30x~30 meter square. There are probably features of interest that are small enough to hide within the large pixels. We believe that a more appropriate resolution is on the order of 5 meters.

## Appendix A.2 Asymmetric Warfare Events

Asymmetric warfare events were obtained from the ISAF-NATO Civilian Integration Team. The events are provided as a service to contractors and non-governmental agencies that operate or may operate in Afghanistan. An UNCLASSIFIED//FOUO extract of the Afghanistan SIGACTS database, this dataset consists of a variety of events including IEDs, direct fire, indirect fire, surface-to-air fire and more. The data is provided with at least one week delay and occasional outages and missing data occur.

The dataset used for this analysis includes 33,140 events that occurred between February 01, 2011 and August 23, 2012. Of these, 13,295 were classified as IED and 16610 were classified as direct fire.

Table 10 shows the dates covered by the current dataset.

**Table 10. Date coverage of asymmetric warfare events in the ISAF-NATO Civilian Integration Team Unclassified Dataset.**

| | 2011 | | | | | | | | | | | | 2012 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Day of the month | J | F | M | A | M | J | J | A | S | O | N | D | J | F | M | A | M | J | J | A |
| | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 |
| | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 |
| | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 |
| | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 |
| | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 |
| | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 | 27 |
| | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 | 28 |
| | 29 | | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 | 29 |
| | 30 | | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | | 30 | 30 | 30 | 30 | 30 | 30 |
| | 31 | | 31 | | 31 | | 31 | 31 | | 31 | | 31 | 31 | | 31 | | 31 | | 31 | 31 |

Data not available from the CIT website.

The following figures offer an overview of the geographic distribution of IED and direct fire events in Afghanistan.

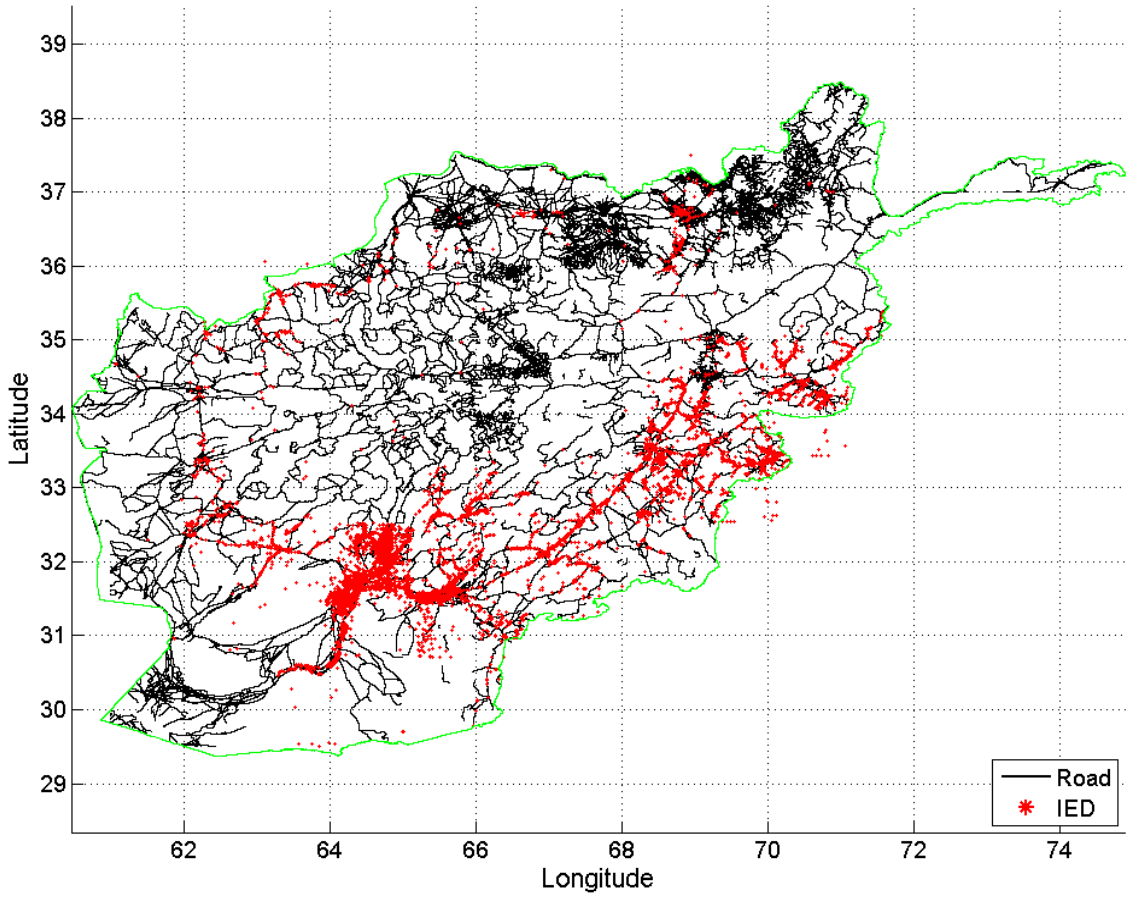



**Figure 67. IED attacks in Afghanistan, February 2011 - August 2012.**

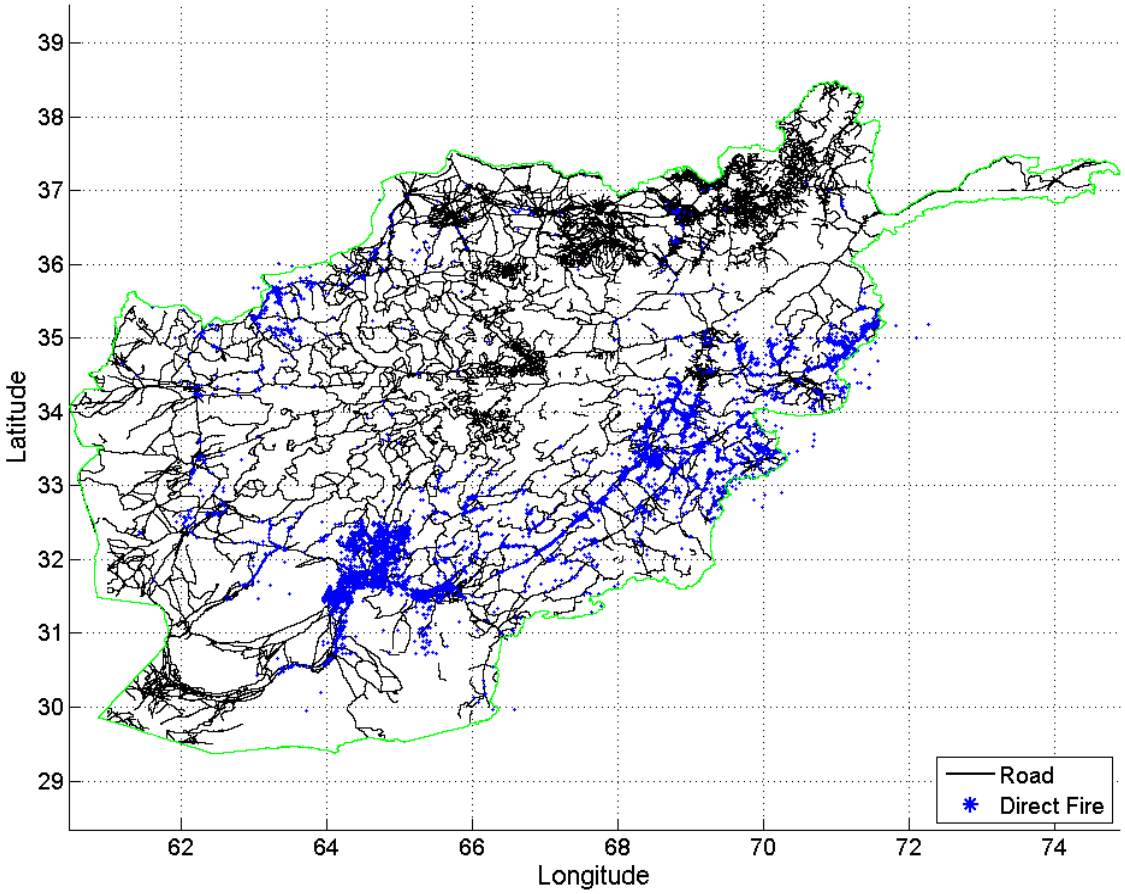



**Figure 68. Direct fire attacks in Afghanistan, February 2011 - August 2012.**

Figure 69 shows the distribution of events by date and type using a 7-day sliding window. All of the events show the same general trends. Note that the trend to zero in August 2012 is an edge effect based on availability of data. The dips in Sep-Oct 2011 and Feb-Mar 2012 are also due to missing data.

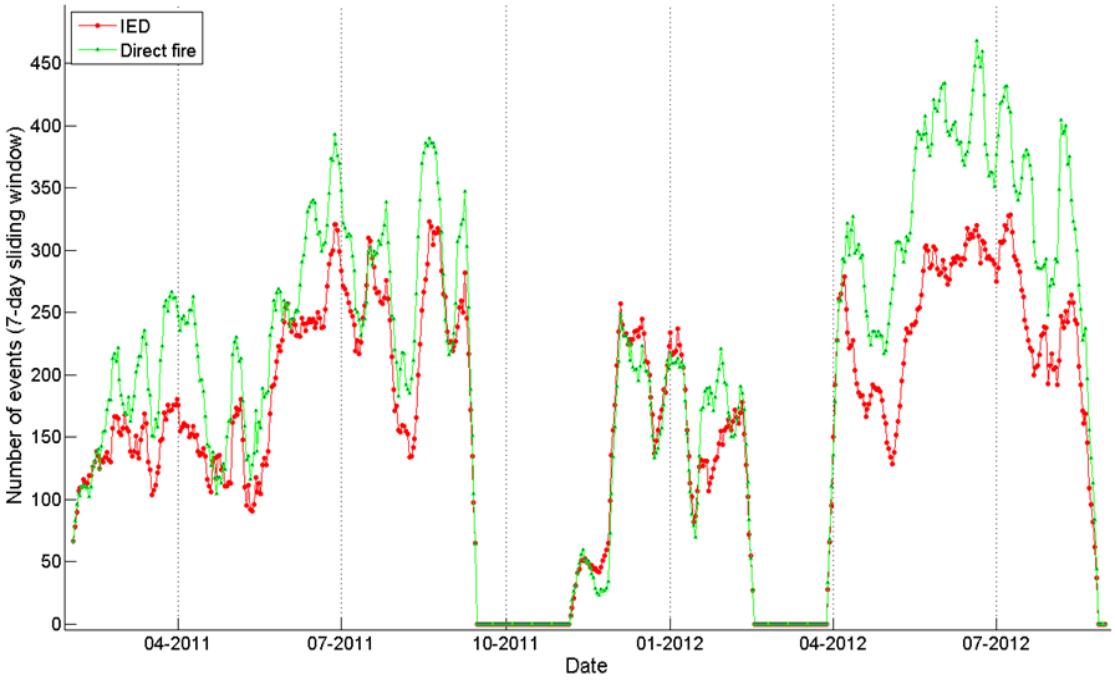



**Figure 69. Distribution of events by date, using a 7-day sliding window.**

*Collocated Events*

Some locations also lend themselves to multiple events, sometimes of the same type and sometimes not. In the following analysis, collocated events are defined as successive events that occur within 250 meters and 1.5 hours of each other. The threshold of 250 meters was selected to account for typical patrol configurations, where vehicles are separated by 25 meters or more. The time window was selected based on anecdotal information about typical patrol behavior following an attack. A total of 894 initiating events meet these criteria.

**Table 11. Collocated events**

| Initiating Event | Following Event | Count |
|---|---|---|
| IED | IED | 319 |
| IED | Direct fire | 63 |
| Direct fire | IED | 42 |
| Direct fire | Direct fire | 470 |

*Dataset Problems*

A principle problem with this conflict event dataset is data quality. In particular, the exact coordinates of events seem to be collected in a variety of ways and using a variety of datums. No information is provided to assess or normalize these inputs. Thus, the dataset is likely to contain locations that are erroneous due to estimation errors, datum translation errors, and simple manual data entry error.

Another problem is the lack of descriptor specificity. All IED events are classified with the same descriptor. Thus, command-detonated and victim-detonated devices are labeled with the same descriptor. Similarly, all direct fire events share the same label. So, a company-sized ambush and an individual sniper attack are marked as being part of the same class.

A final problem is the lack of consistency in the measurements. For an ambush, the specified location is likely to be the geographic coordinate of the person reporting the event when it started. Depending on the size of the convoy or patrol, this location could be tens to hundreds of meters from the actual place where the attack actually occurred. If the patrol was moving during the attack, the location may be estimated or may be the place where the patrol stopped. Similar problems exist for IED events.

## Appendix A.3 Afghanistan Roads

Road data is collected and maintained by the Afghanistan Information Management Service (http://www.aims.org.af) and distributed by mapcruzin.com at http://www.mapcruzin.com/afghanistan-shapefiles/roads.zip. Three types of roads are identified, including all weather primary, all weather secondary and tracks.

The roads are stored as polyline objects in a shapefile. In order to use the roads for this project, each road segment was split into discrete points at 30 meter intervals (the resolution of the elevation maps). A total of 3,306,680 discrete points were produced in this way.

Principle problems with this dataset include its age and apparent incompleteness. As far as can be determined, this map of roads was produced in the early 2000's using data from Russian and U.S. maps published in the 1980's. The age of the data suggests that some current roads may be missing from the map, particularly after post-war reconstruction efforts by the U.S. and others. Figure 70 illustrates the problem. A red square draws attention to a number of IED events that seemed to occur away from roads. Visual analysis of Google Earth imagery reveals the presence of a road and a number of villages along a watercourse. It seems likely that this road did not exist or was not surveyed when the Russian and U.S. maps were originally created.

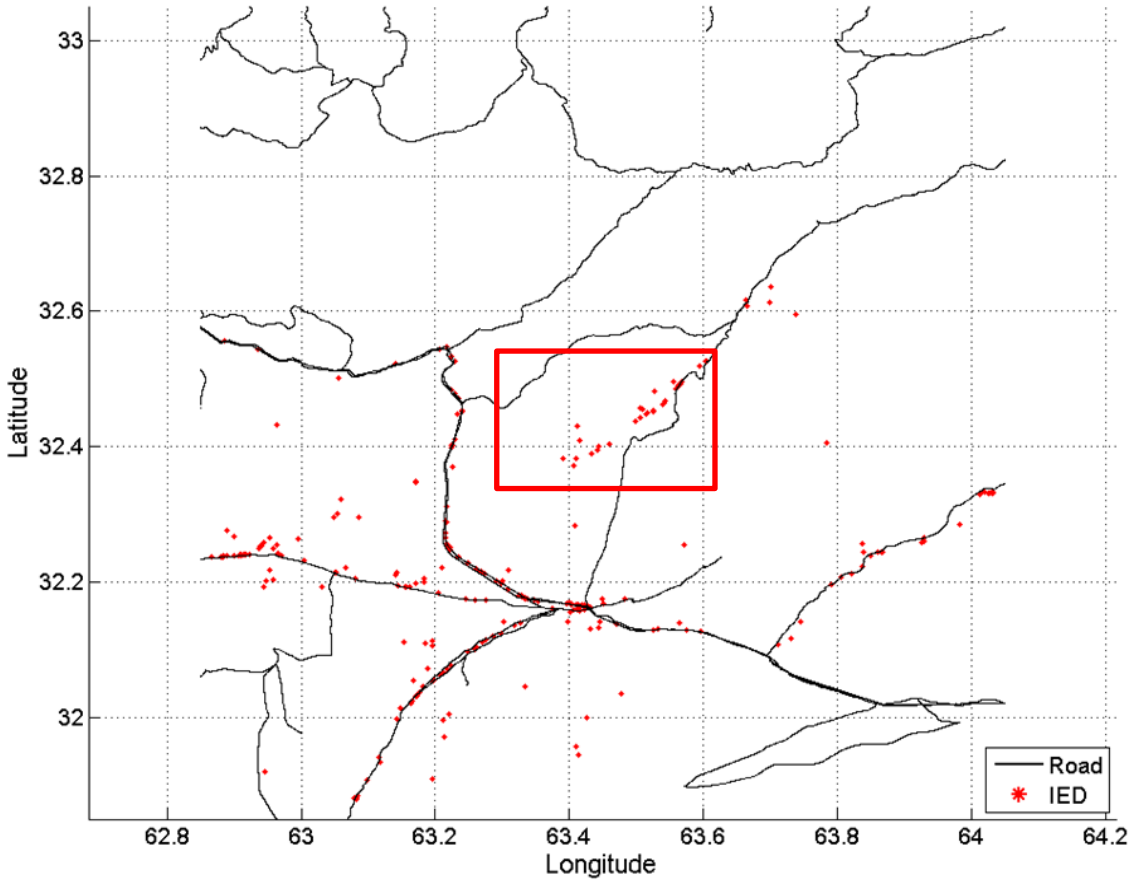



**Figure 70. Example of area where IED events occurred away from known roads.**

## Appendix A.4 Population

Population estimates were scraped from the site http://www.fallingrain.com/world/AF/. For each known, fixed, and named populated place (village, town, city) the total population within 7 km is estimated. Figure 71 gives an idea of the population distribution in Afghanistan.

The source of the raw population data is unknown. Therefore, the validity of the estimated populations is unknown. Additionally, as a largely rural and tribal society, Afghan participation in a national census is likely to be less than complete. Informally, the fallingrain.com estimates for the area surrounding the center of Kabul, Kandahar and Mazar-i-sharif appear to be roughly consistent with Wikipedia estimates for population for the same towns.

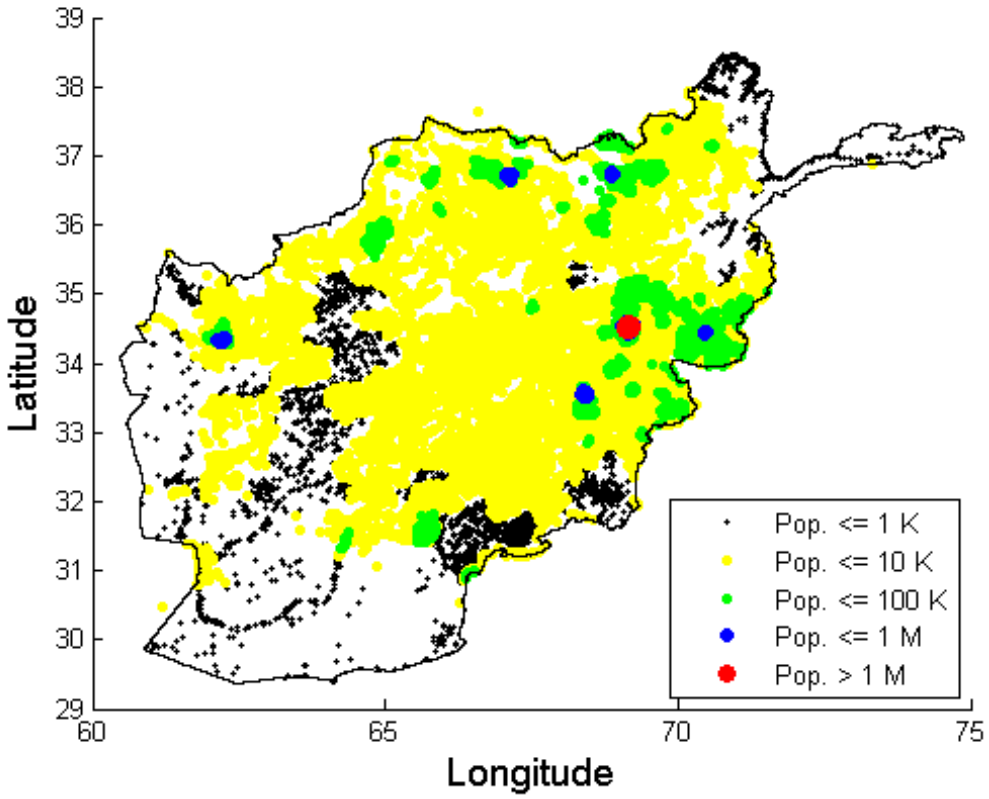



**Figure 71. Estimated population at locations throughout Afghanistan.**

APPENDIX B. ADDITIONAL FEATURES

This appendix contains additional features. For the most part, these features are similar to those already presented in the body of the dissertation and differ only in window size, radius, or number of radials. They are included here for completeness.

## Appendix B.1 Visibility Index

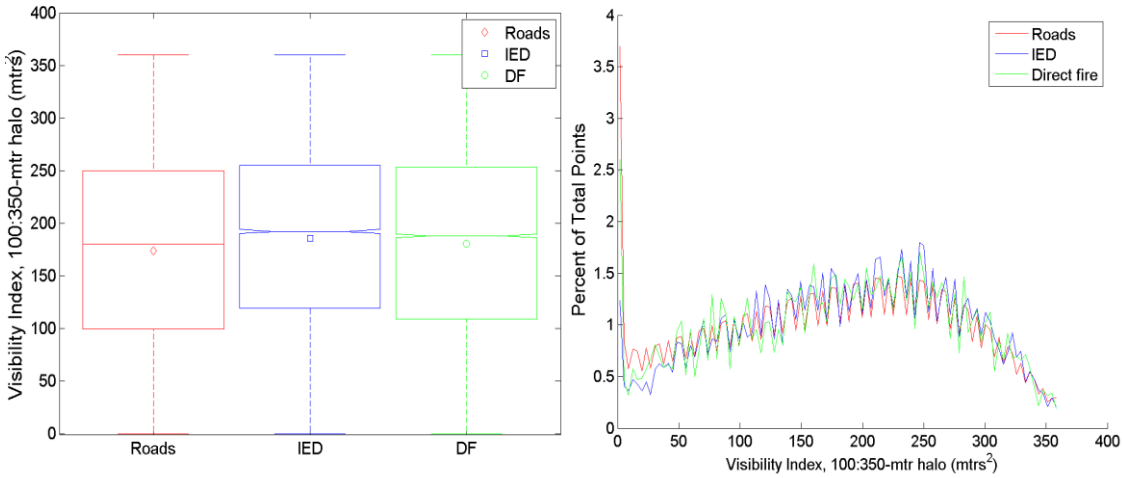



**Figure 72. Visibility Index inside a halo with an inner radius of 100 meters and an outer radius of 350 meters.**

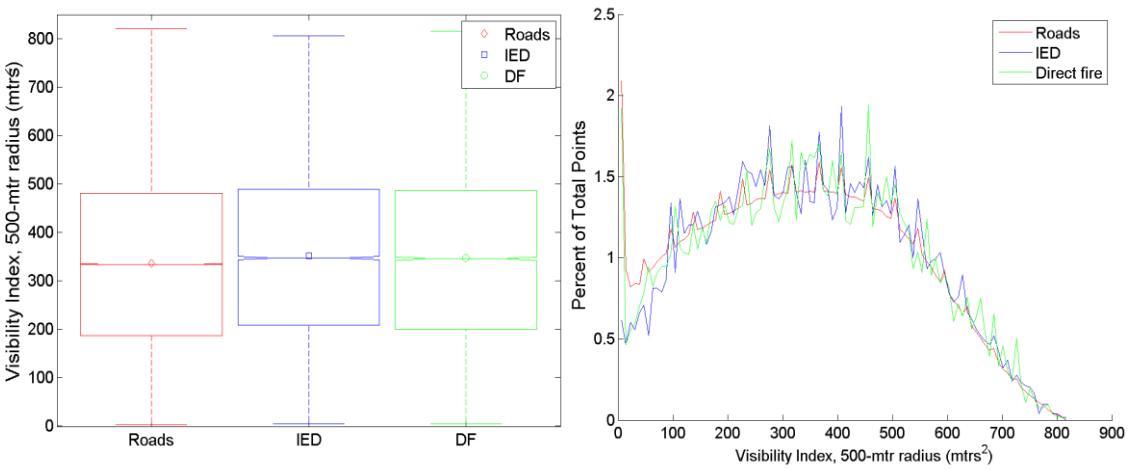



**Figure 73. Visibility Index at a radius of 500 meters.**

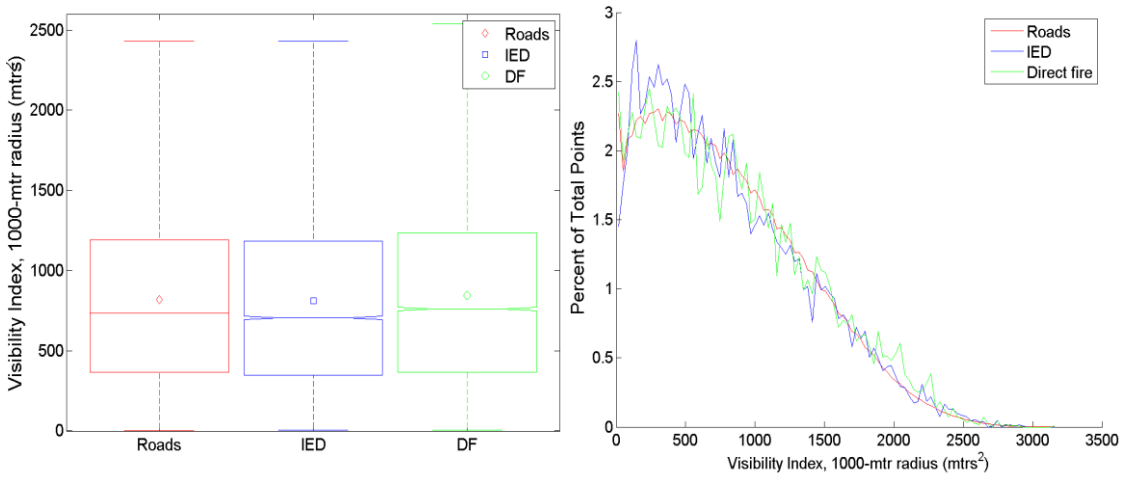



**Figure 74, Visibility Index at a radius of 1000 meters.**

## Appendix B.2 Discrete Shape Complexity Index

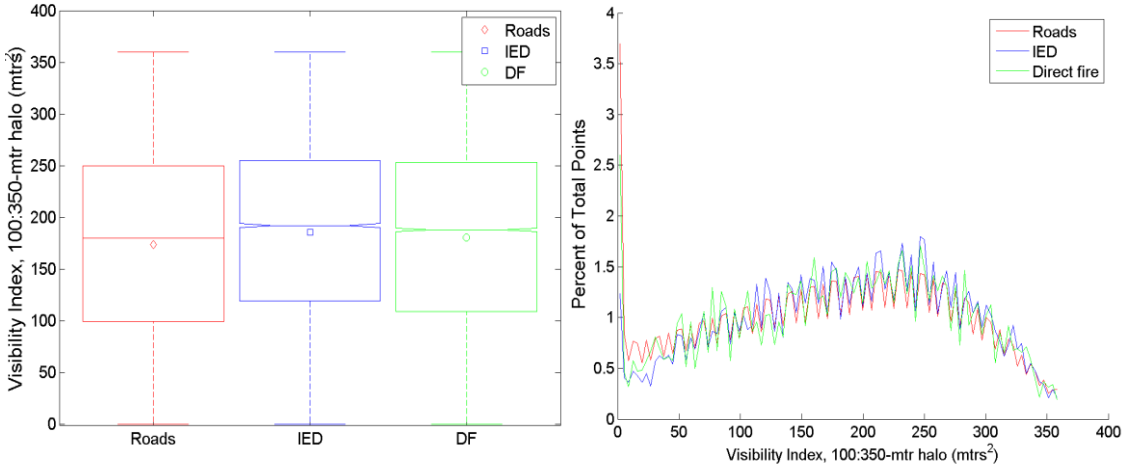



**Figure 75. Discrete Shape Complexity Index in a halo with an inner radius of 100 meters and an outer radius of 350 meters.**

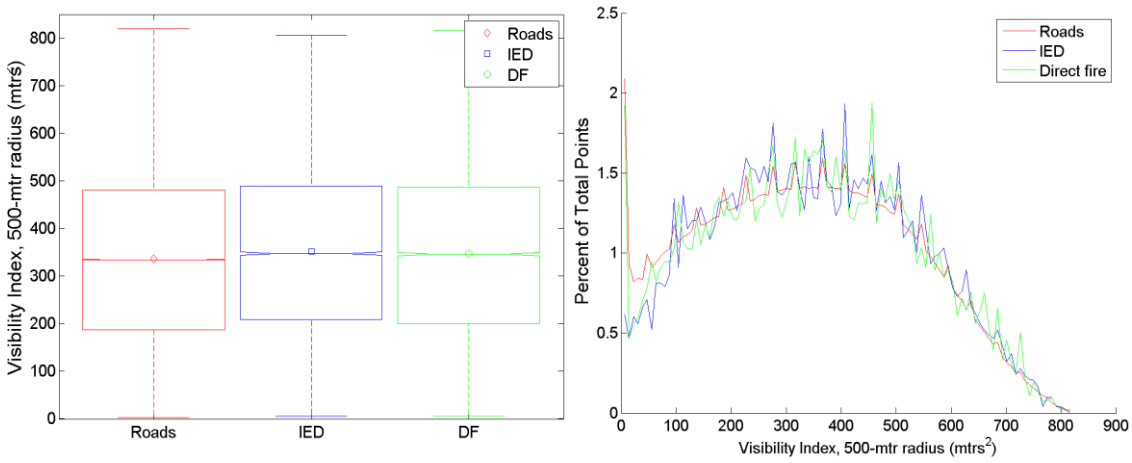



**Figure 76. Discrete Shape Complexity Index at a radius of 500 meters.**

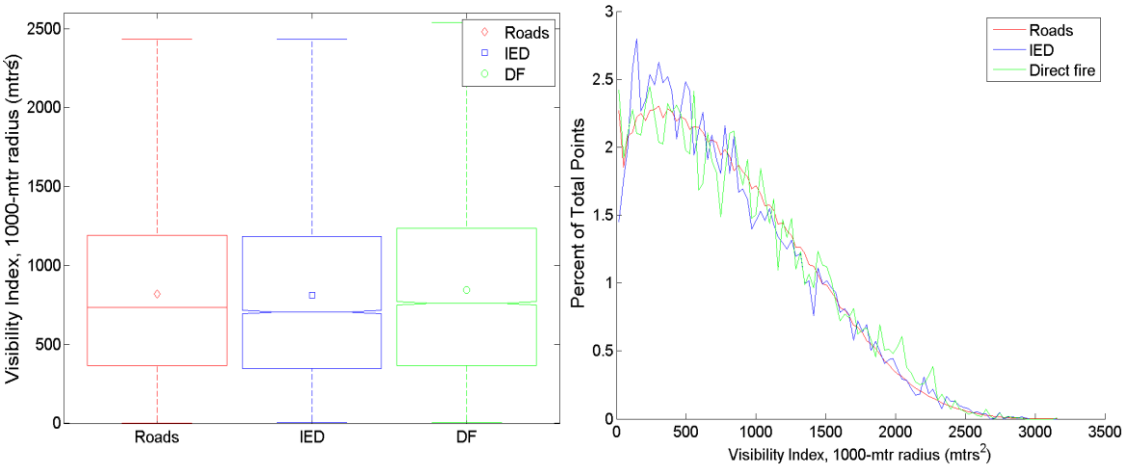



**Figure 77. Discrete Shape Complexity Index at a radius of 1000 meters.**

## Appendix B.3 Cumulative Escape Adjacency for a Single Point

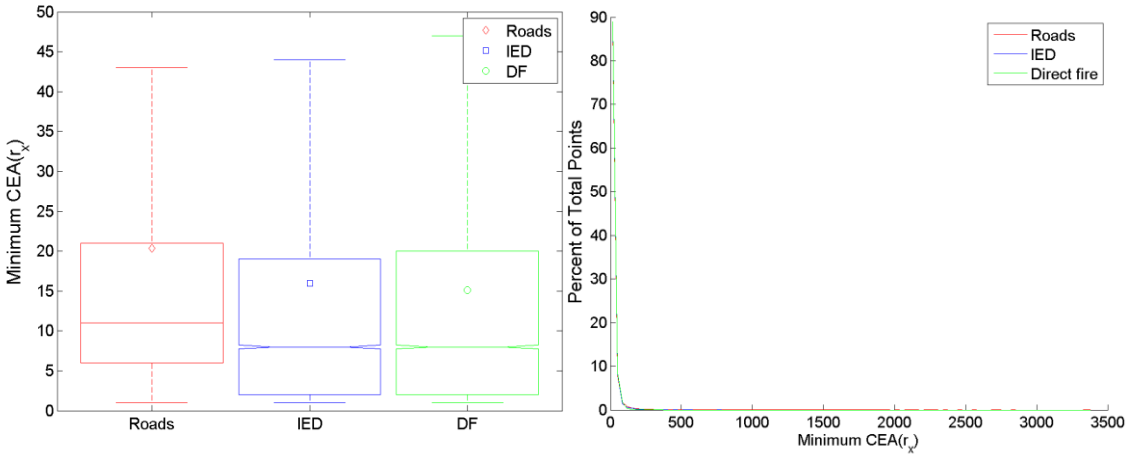



**Figure 78. Minimum CEA($r_x$).**

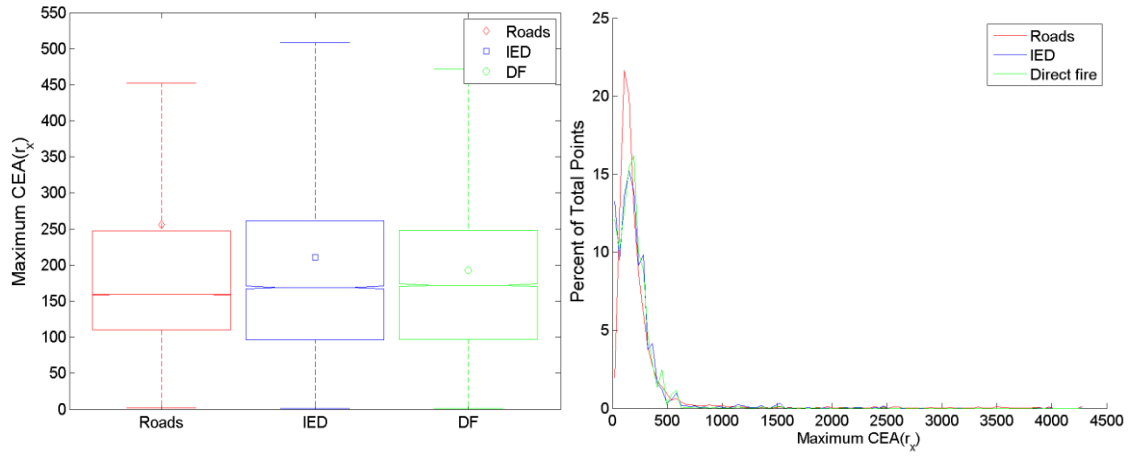



**Figure 79. Maximum CEA($r_x$).**

## Appendix B.4 Median Route Visibility

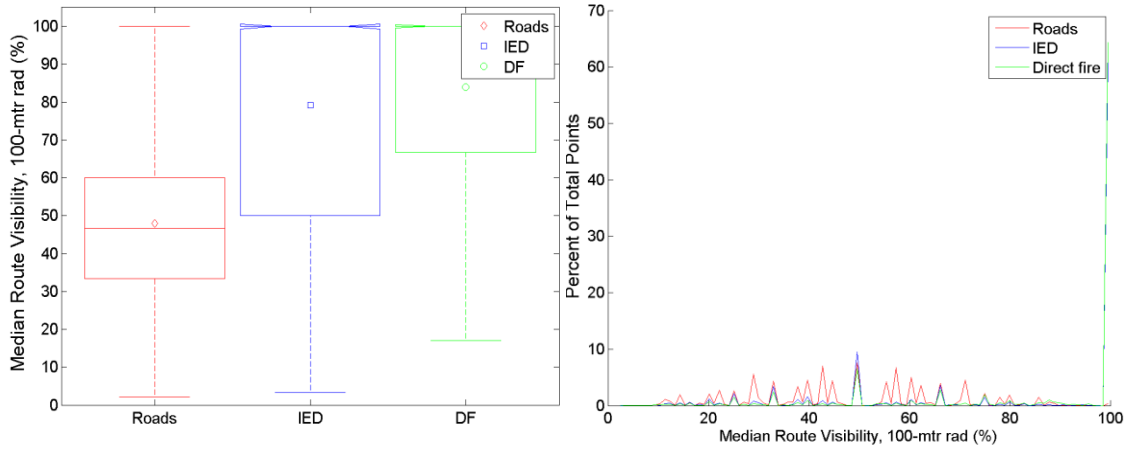



**Figure 80. Median route visibility at 100 meters.**

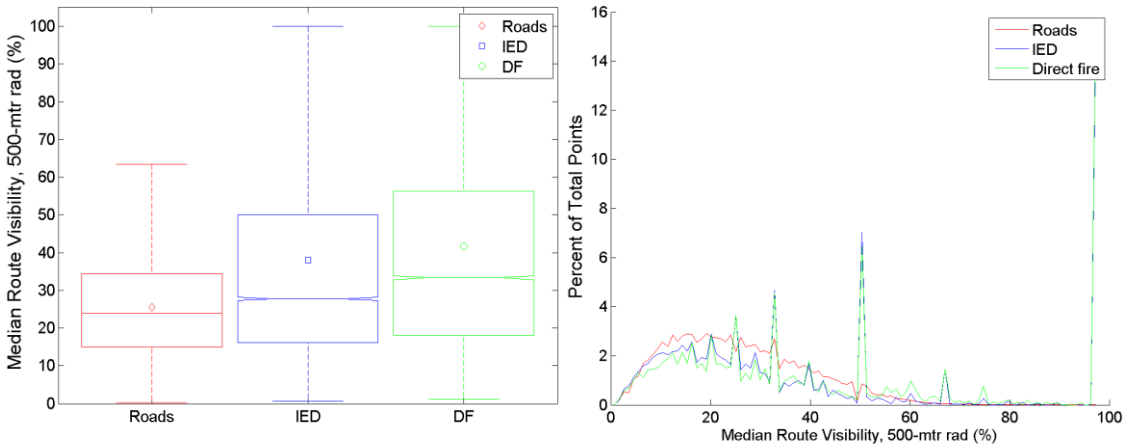



**Figure 81. Median route visibility at 500 meters.**

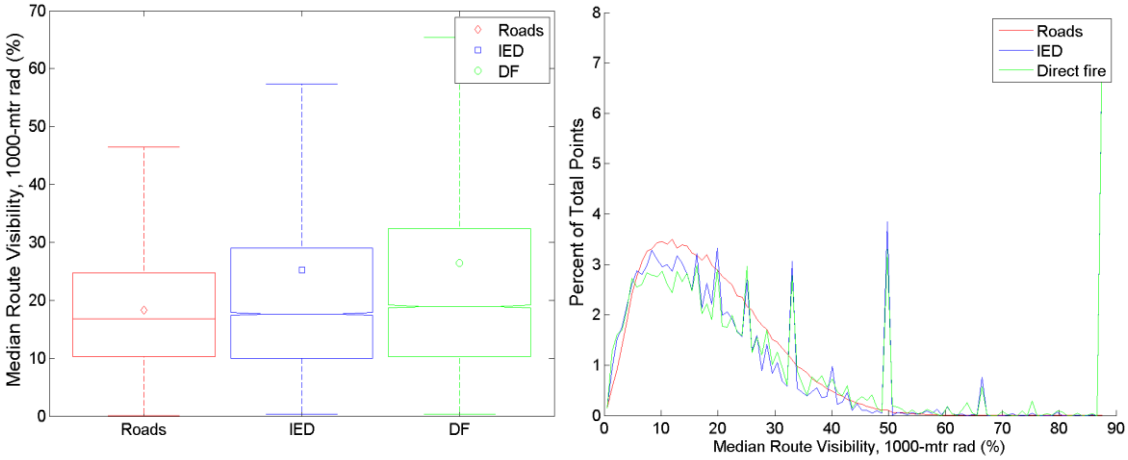



**Figure 82. Median route visibility at 1000 meters.**

## Appendix B.5 Minimum Route Visibility

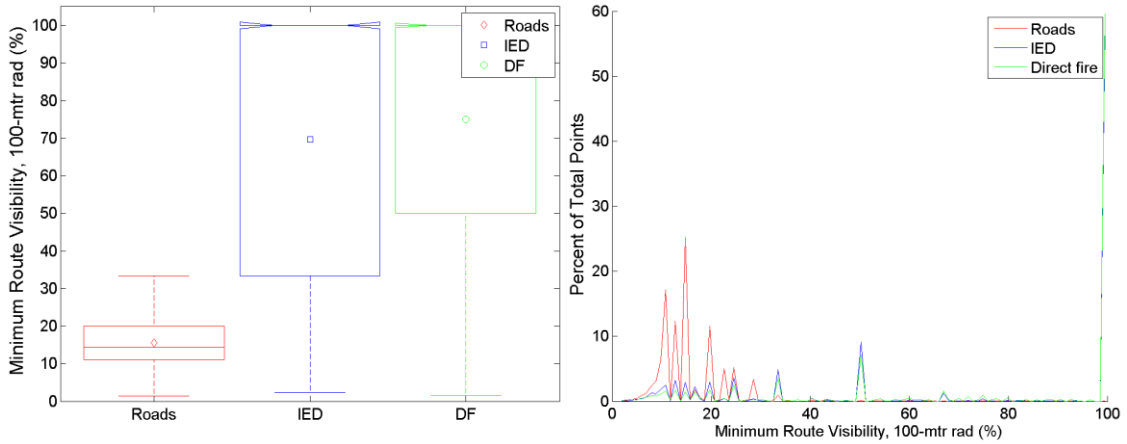



**Figure 83. Minimum route visibility at 100 meters.**

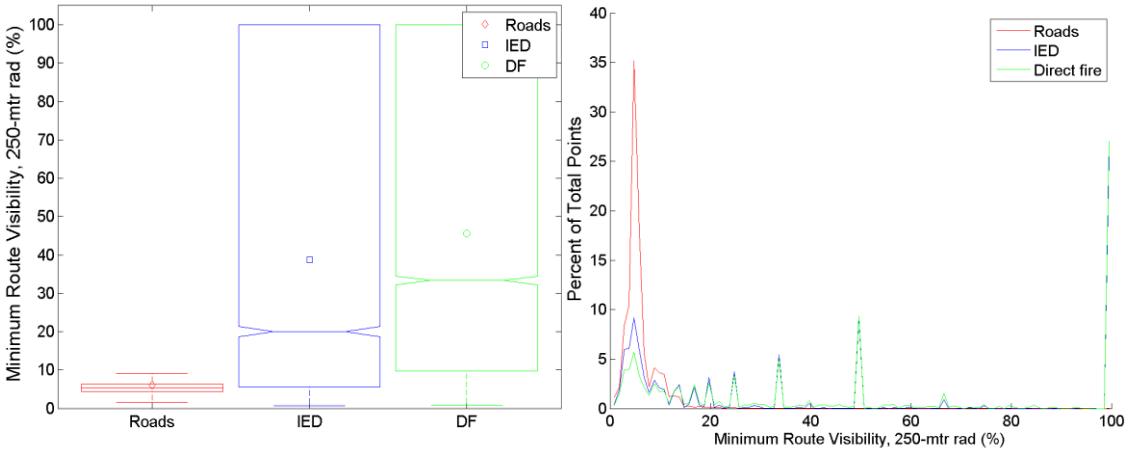



**Figure 84. Minimum route visibility at 250 meters.**

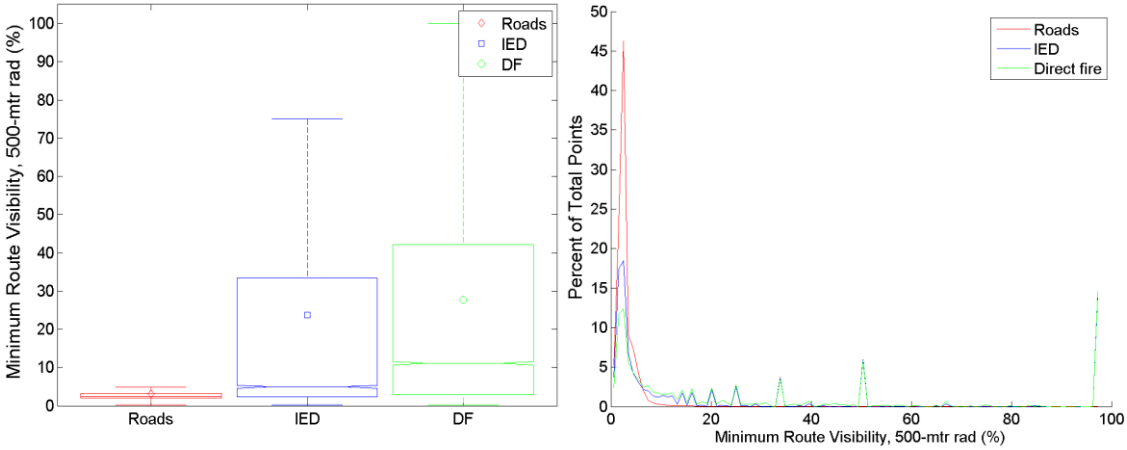



**Figure 85. Minimum route visibility at 500 meters.**

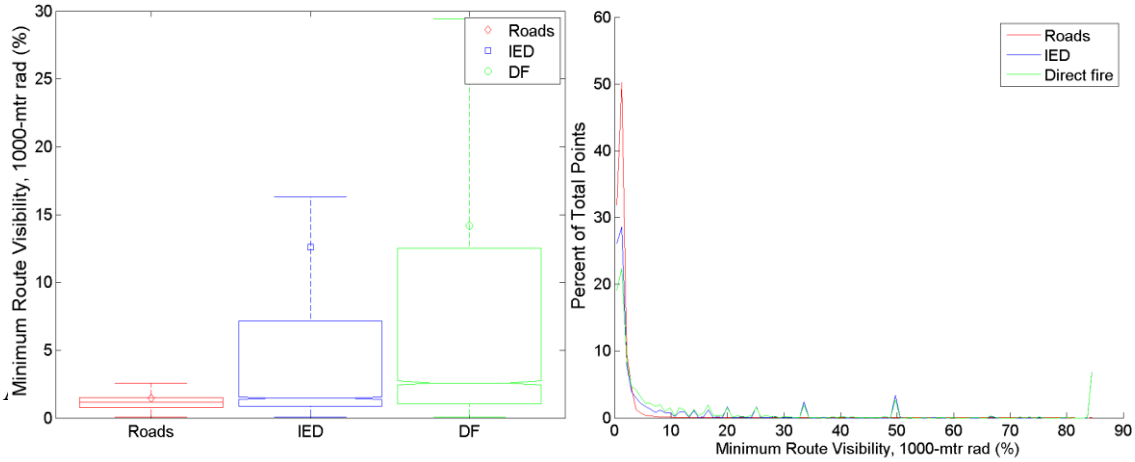



**Figure 86. Minimum route visibility at 1000 meters.**

## Appendix B.6 Maximum Route Visibility

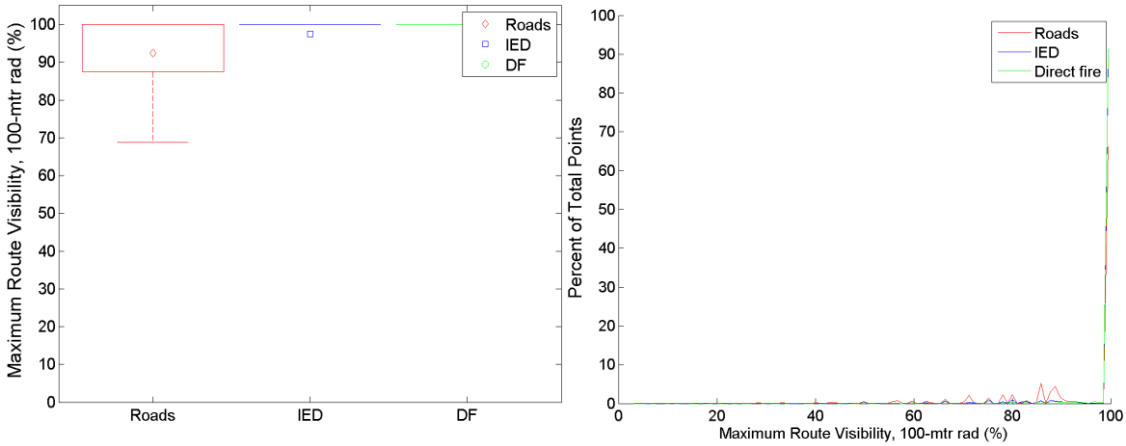



**Figure 87. Maximum route visibility at 100 meters.**

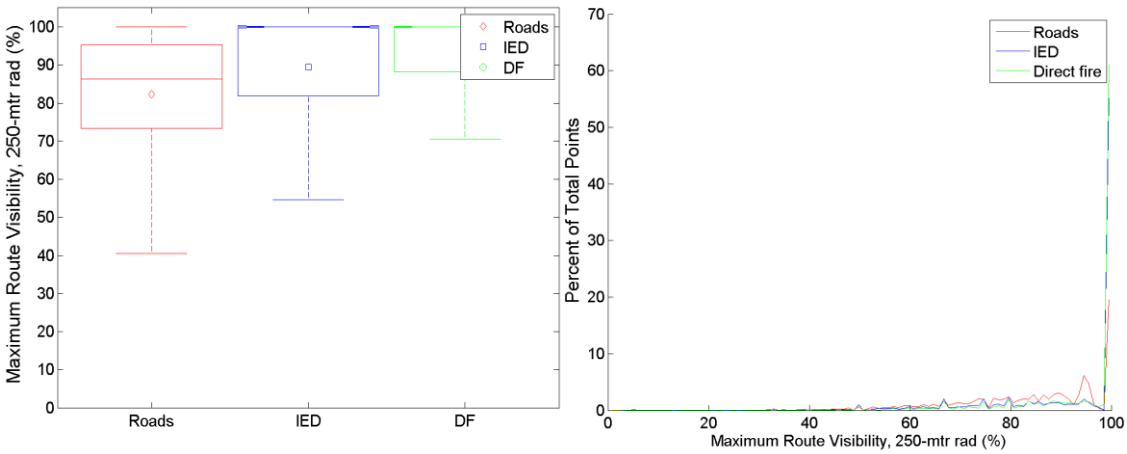



**Figure 88. Maximum route visibility at 250 meters.**

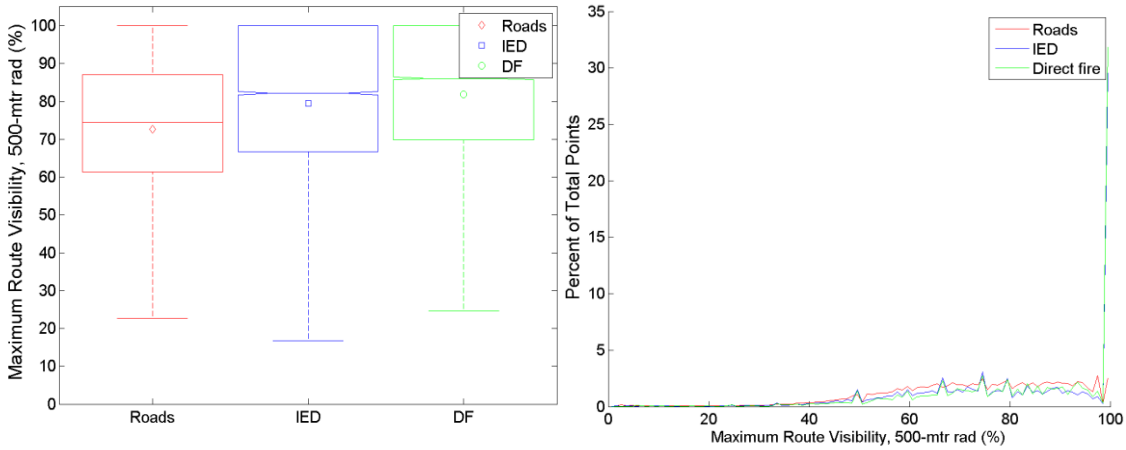



**Figure 89. Maximum route visibility at 500 meters.**

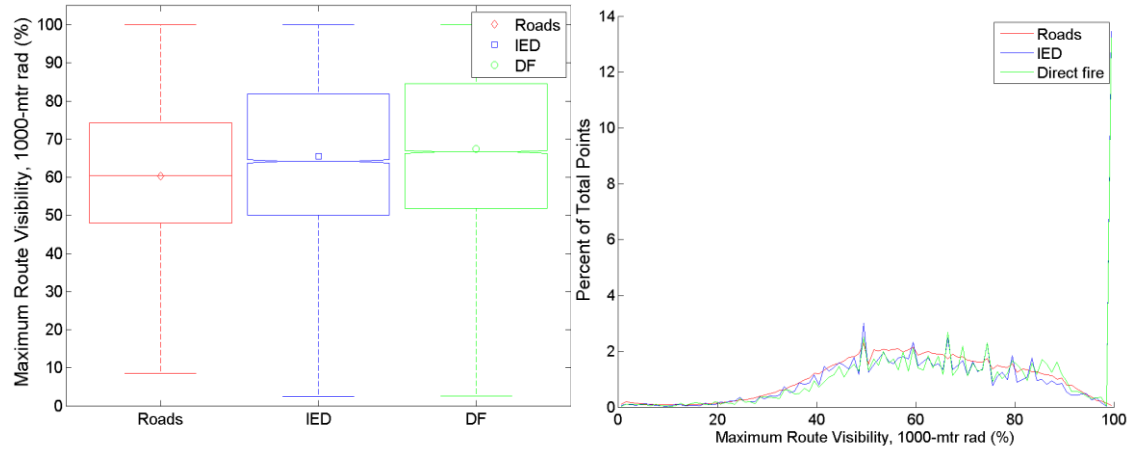



**Figure 90. Maximum route visibility at 1000 meters.**

## Appendix B.7 Sparse Viewshed Shortest Radial

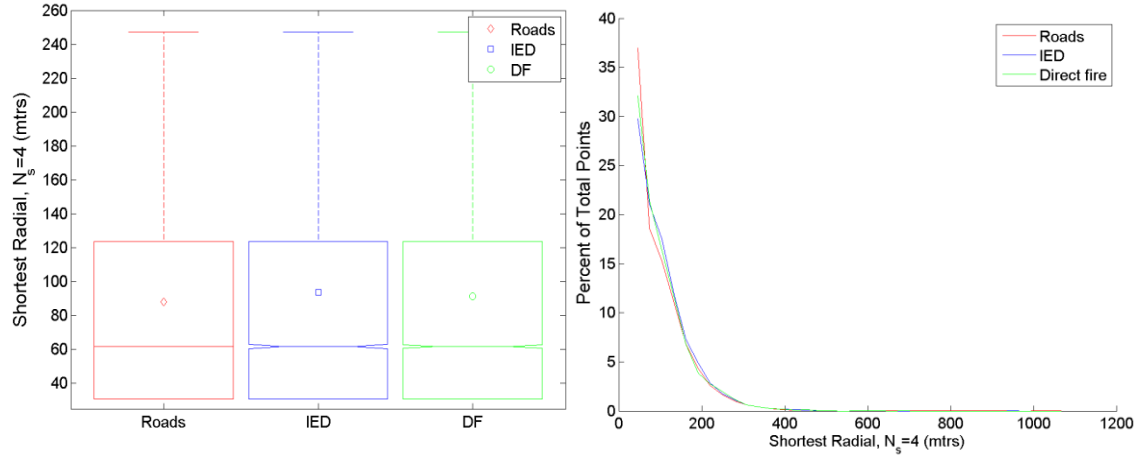



**Figure 91. Sparse viewshed shortest radial ($N_s$ = 4).**

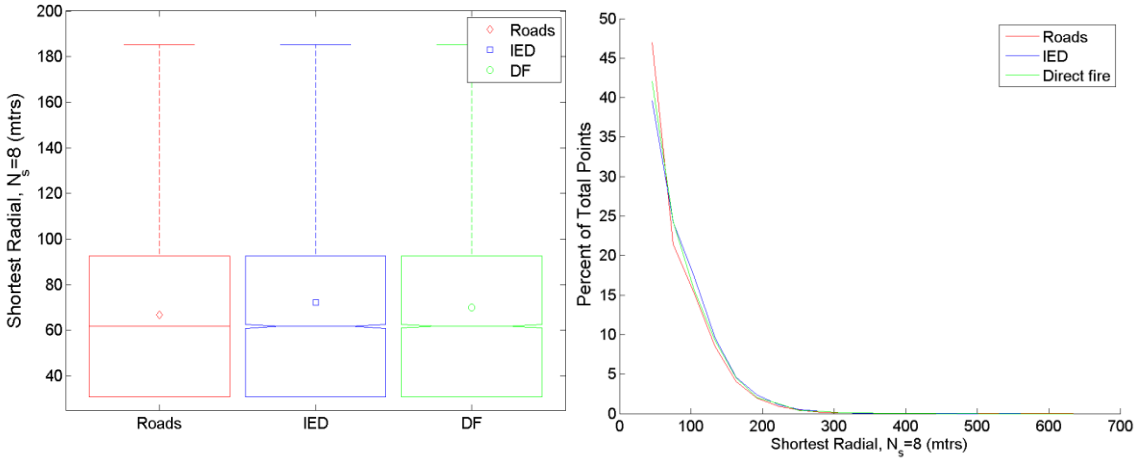



**Figure 92. Sparse viewshed shortest radial ($N_s$ = 8).**

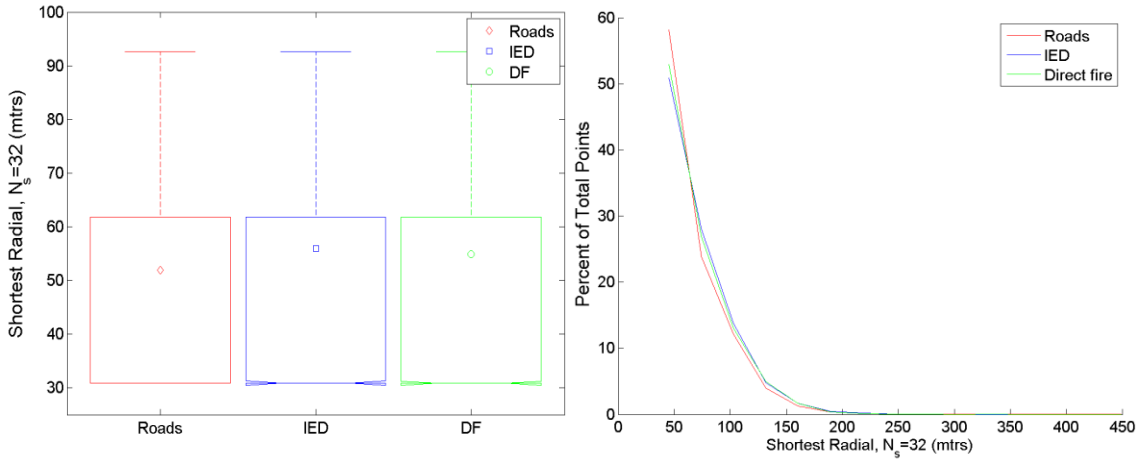



**Figure 93. Sparse viewshed shortest radial ($N_s$ = 32).**

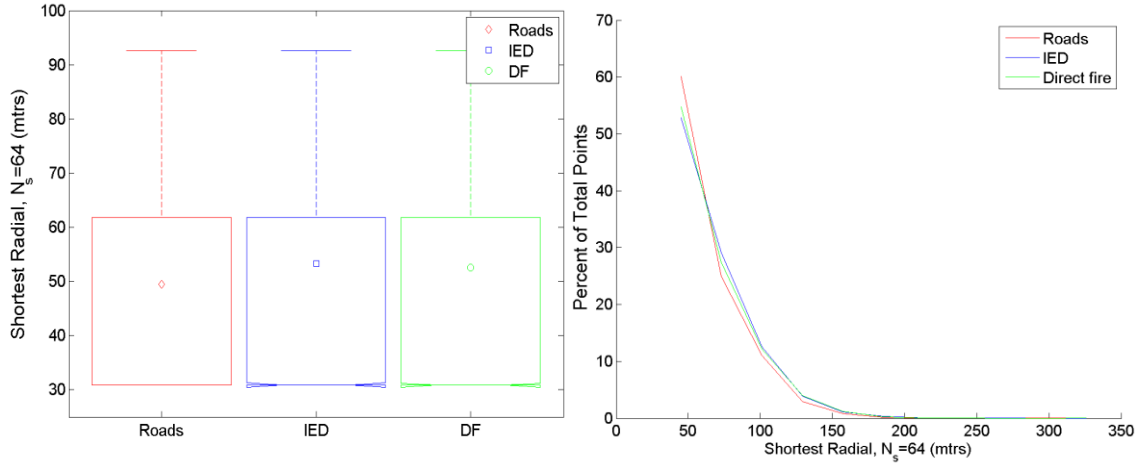



**Figure 94. Sparse viewshed shortest radial ($N_s$ = 64).**

## Appendix B.8 Sparse Viewshed Longest Radial

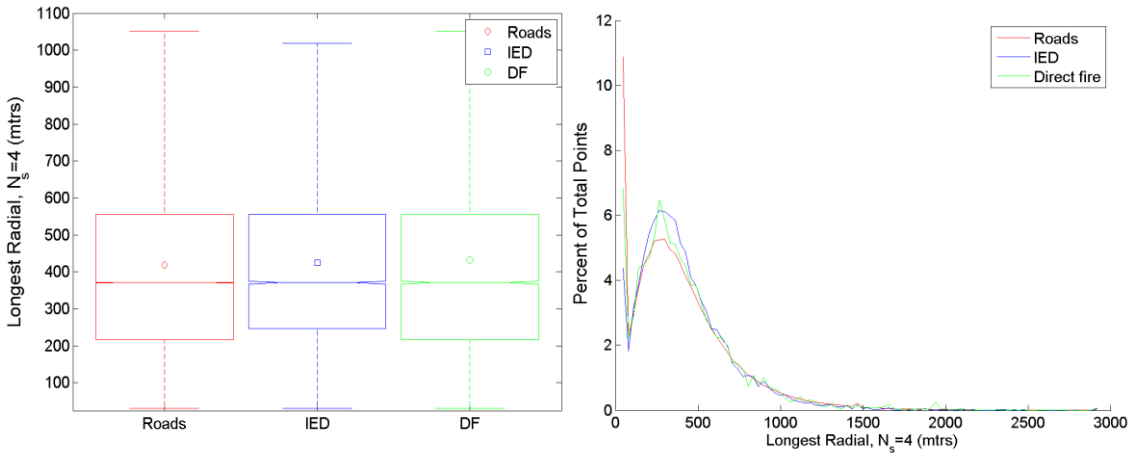



**Figure 95. Sparse viewshed longest radial ($N_s$ = 4).**

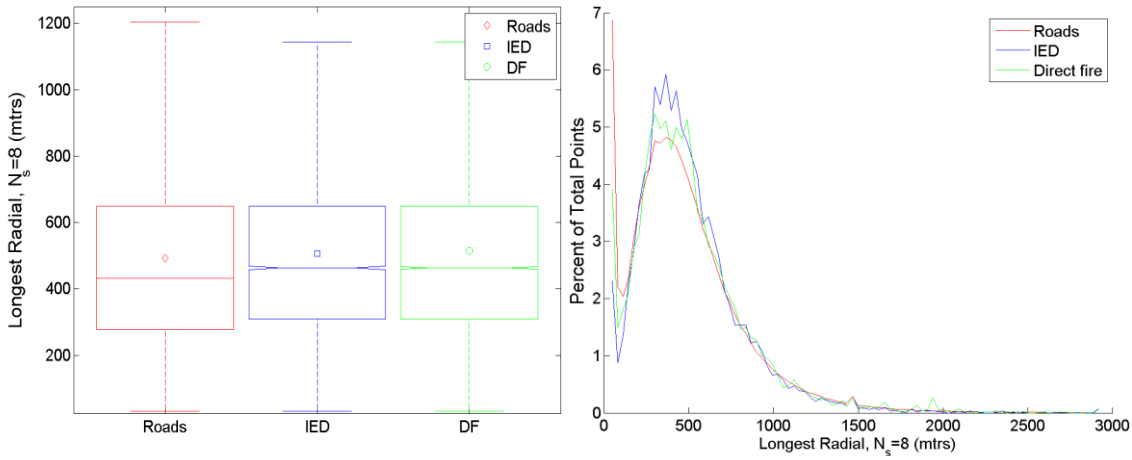



**Figure 96. Sparse viewshed longest radial ($N_s$ = 8).**

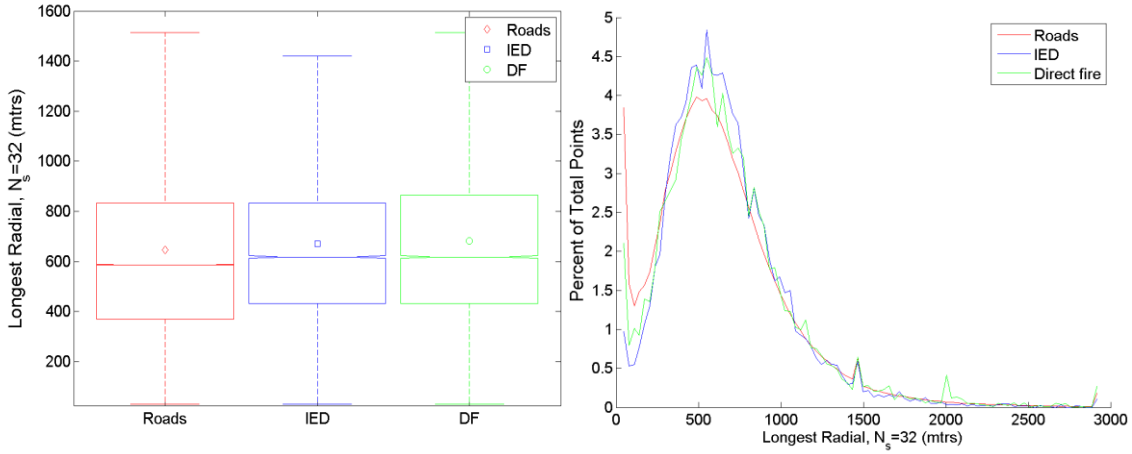



**Figure 97. Sparse viewshed longest radial ($N_s$ = 32).**

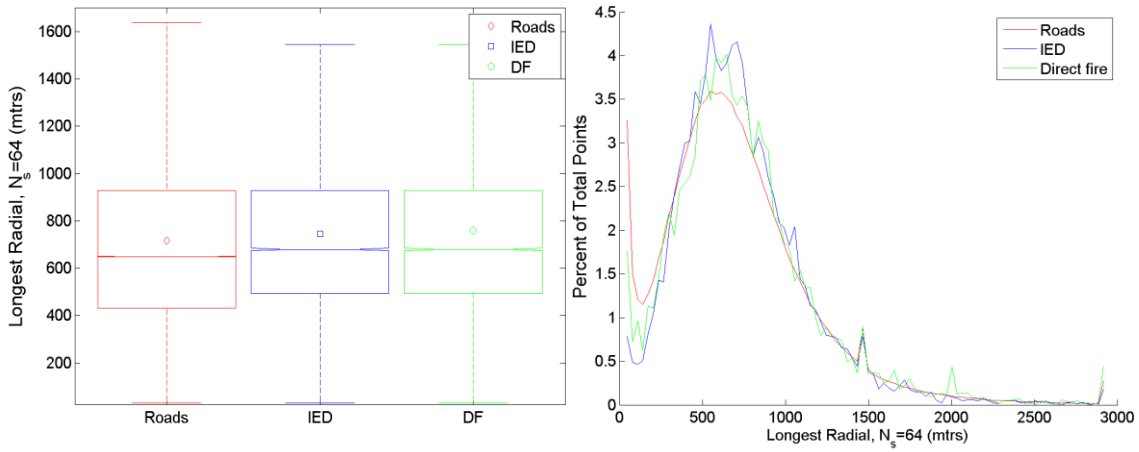



**Figure 98. Sparse viewshed longest radial ($N_s$ = 64).**

## Appendix B.9 Sparse Viewshed Local Openness

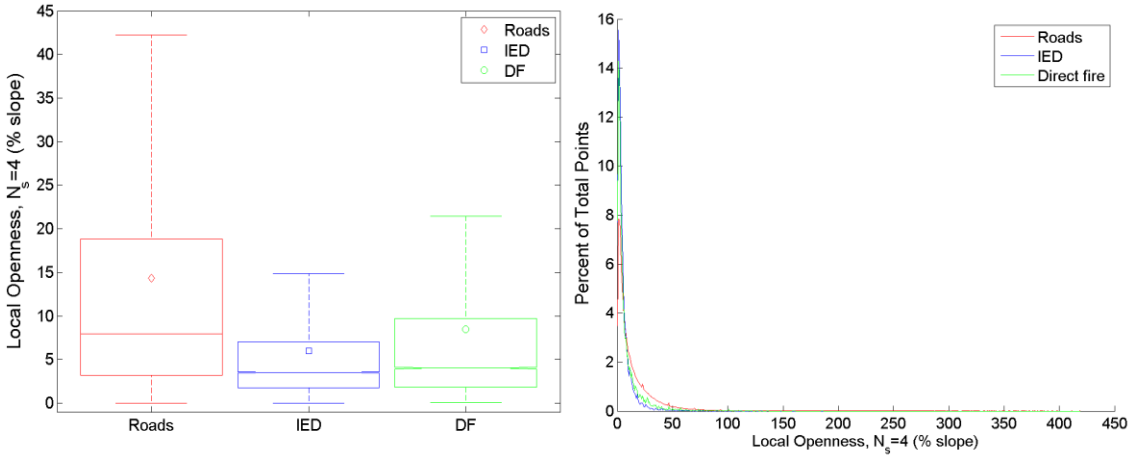



**Figure 99. Sparse viewshed local openness ($N_s$ = 4).**

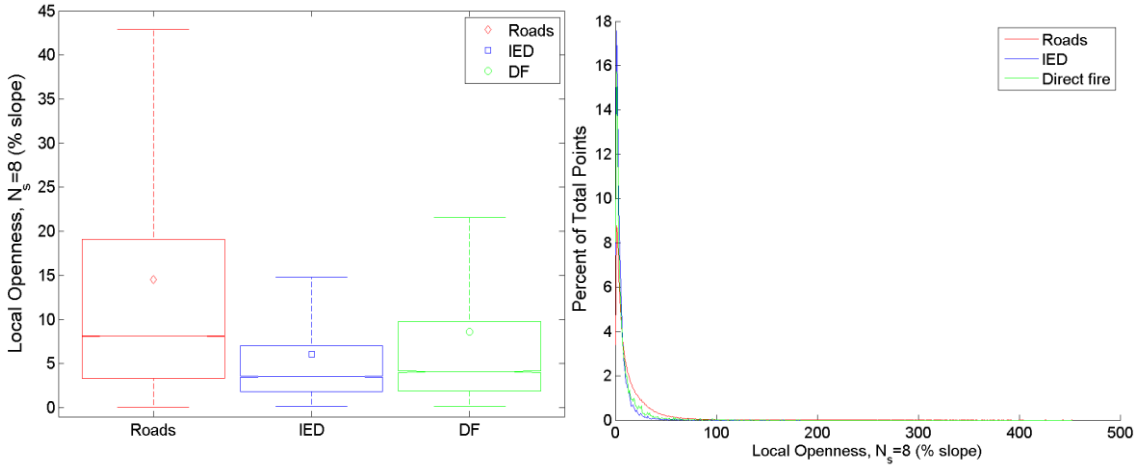



**Figure 100. Sparse viewshed local openness ($N_s$ = 8).**

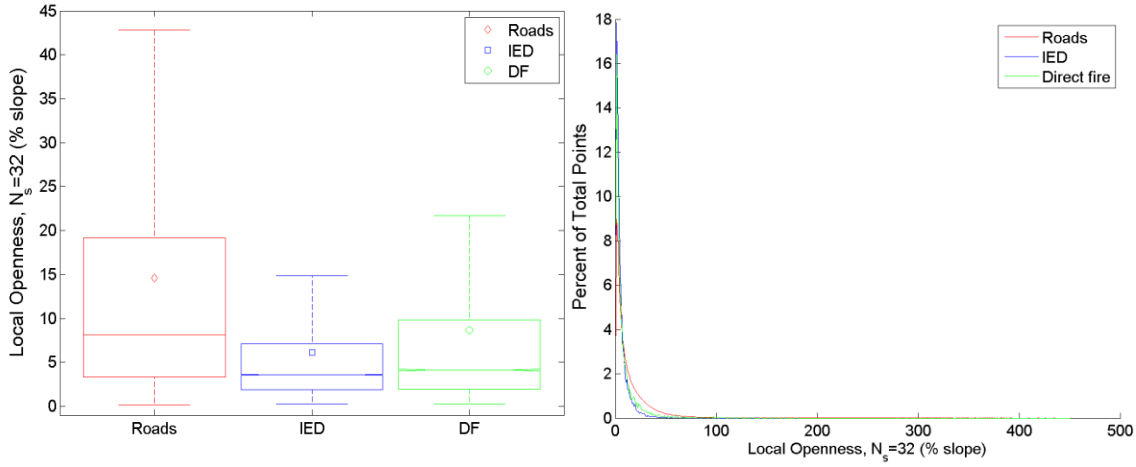



**Figure 101. Sparse viewshed local openness ($N_s$ = 32).**

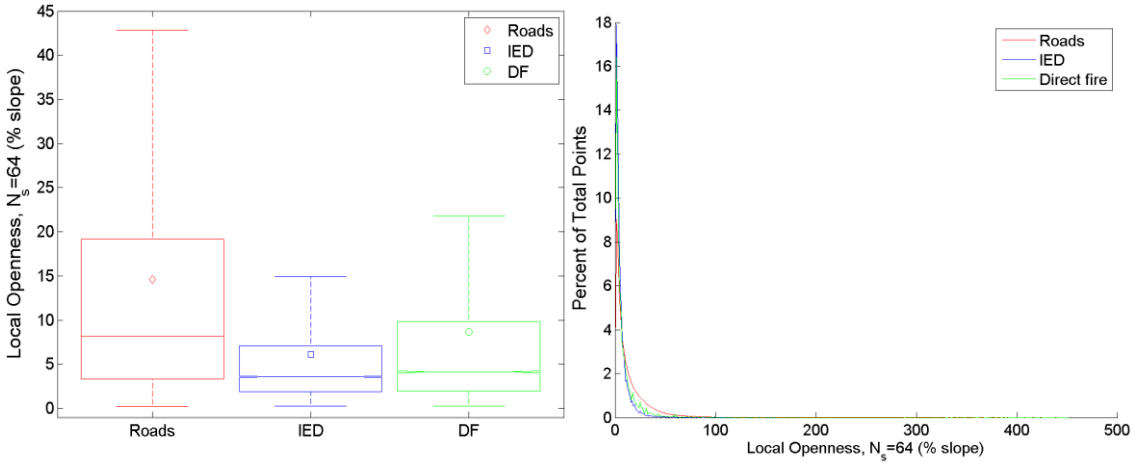



**Figure 102. Sparse viewshed local openness ($N_s$ = 64).**

## Appendix B.10 Elevation Range

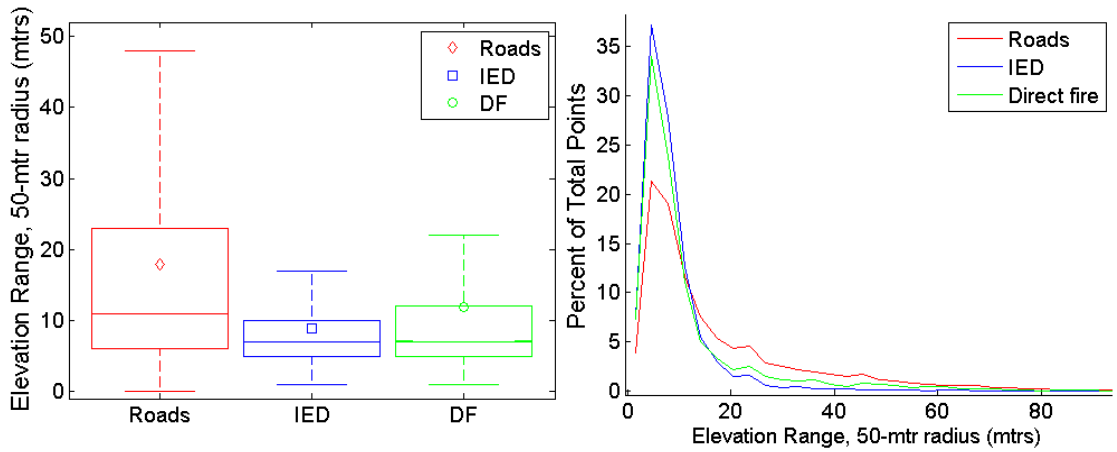



**Figure 103. Elevation range at a radius of 50 meters.**

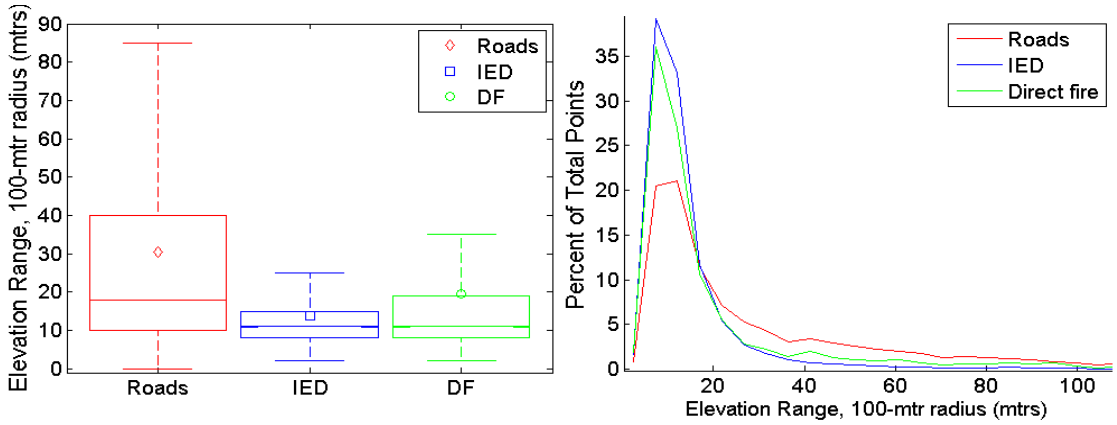



**Figure 104. Elevation range at a radius of 100 meters.**

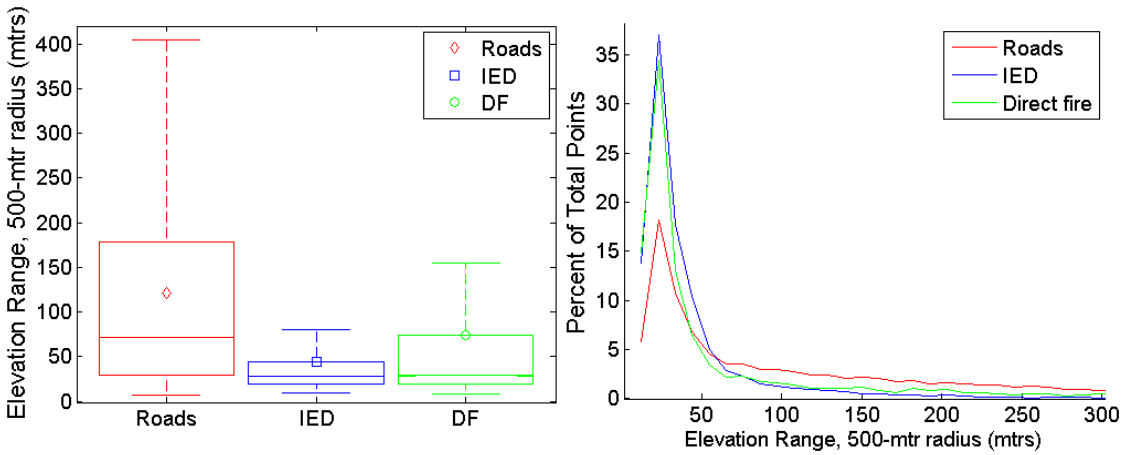



**Figure 105. Elevation range at a radius of 500 meters.**

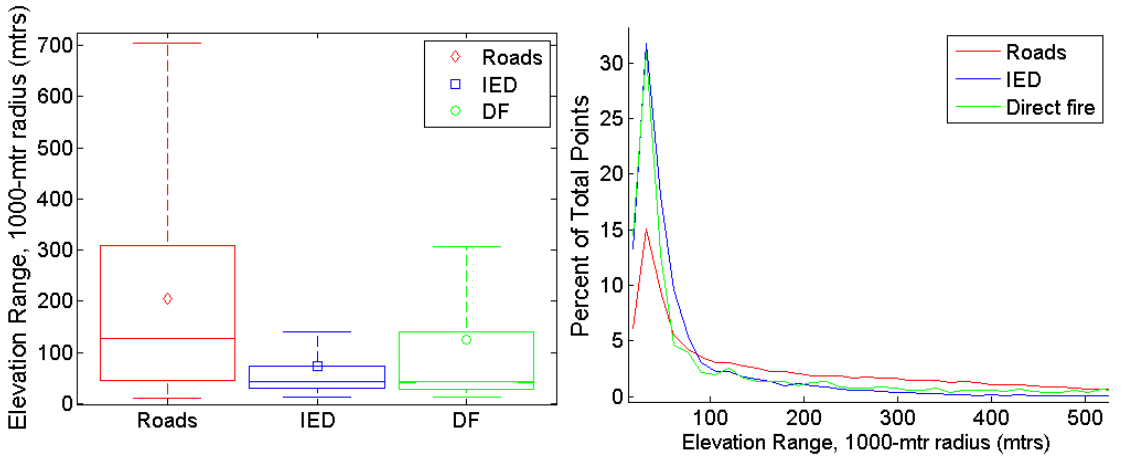



**Figure 106. Elevation range at a radius of 1000 meters.**

## Appendix B.11 Roughness (Standard Deviation of Elevation)

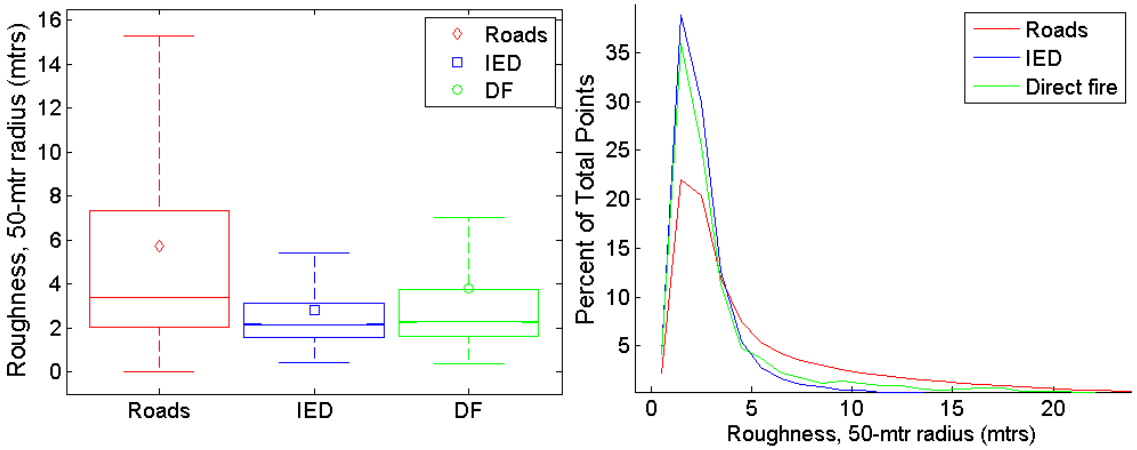



**Figure 107. Roughness at a radius of 50 meters.**

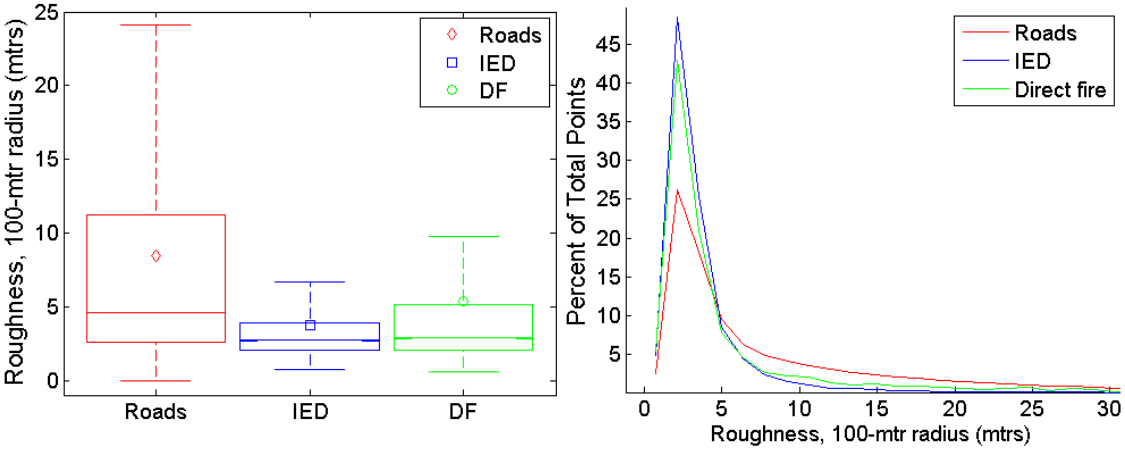



**Figure 108. Roughness at a radius of 100 meters.**

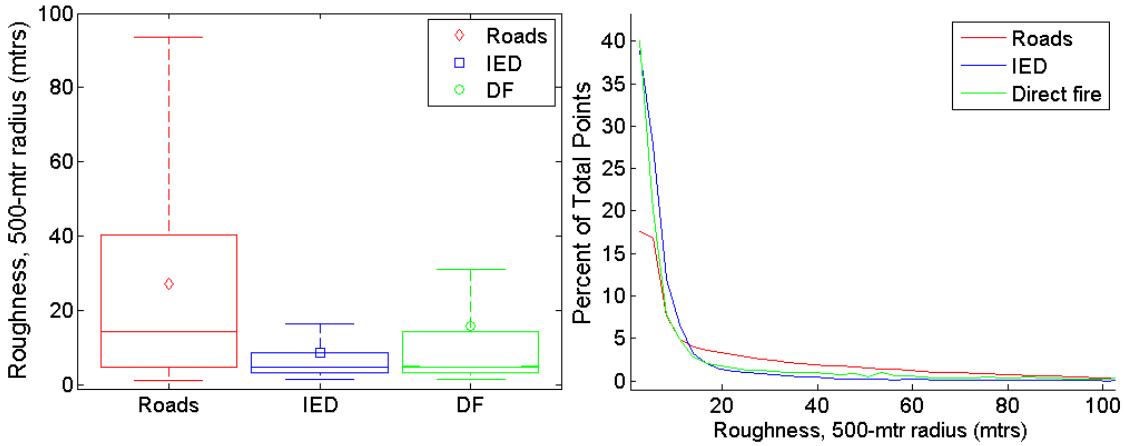



**Figure 109. Roughness at a radius of 500 meters.**

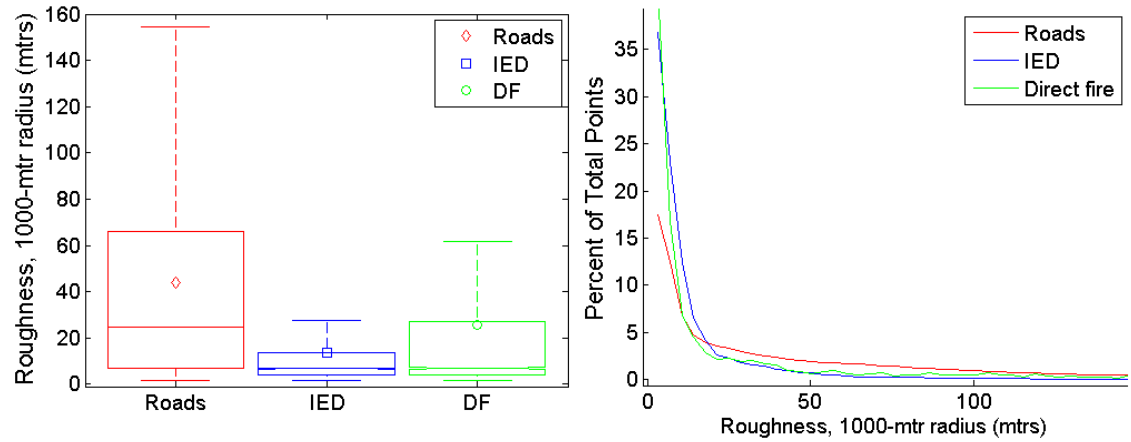



**Figure 110. Roughness at a radius of 1000 meters.**

## Appendix B.12 Sparse Viewshed Mean Radial

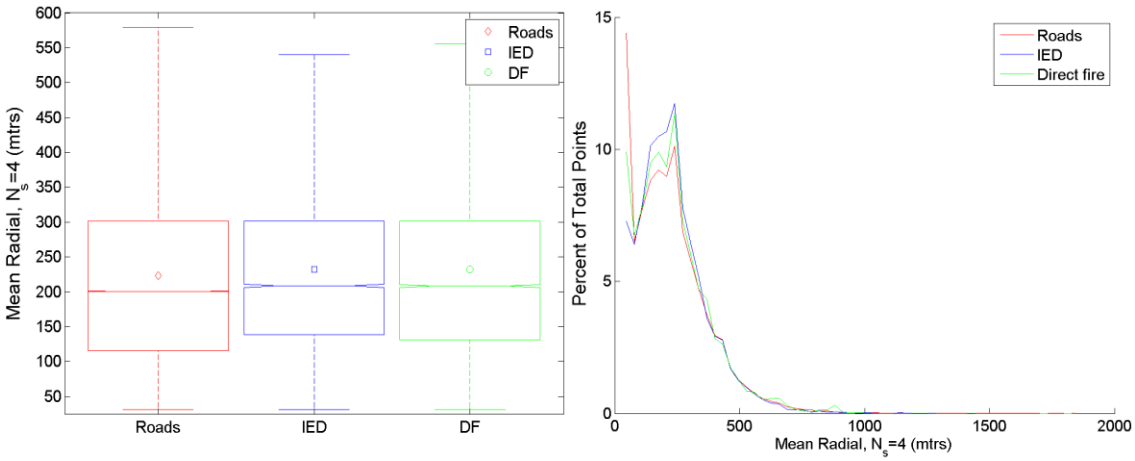



**Figure 111. Sparse viewshed mean radial ($N_s$ = 4).**

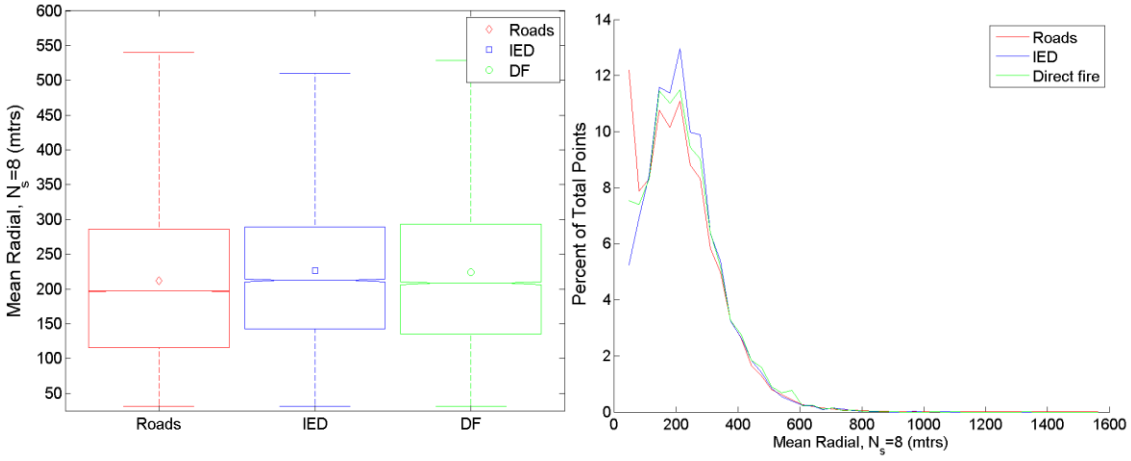



**Figure 112. Sparse viewshed mean radial ($N_s$ = 8).**

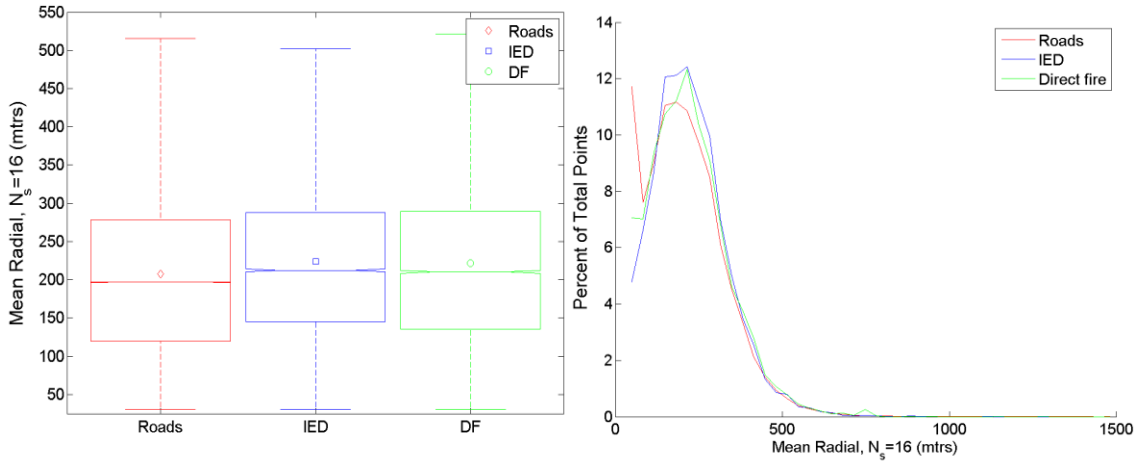



**Figure 113. Sparse viewshed mean radial ($N_s$ = 16).**

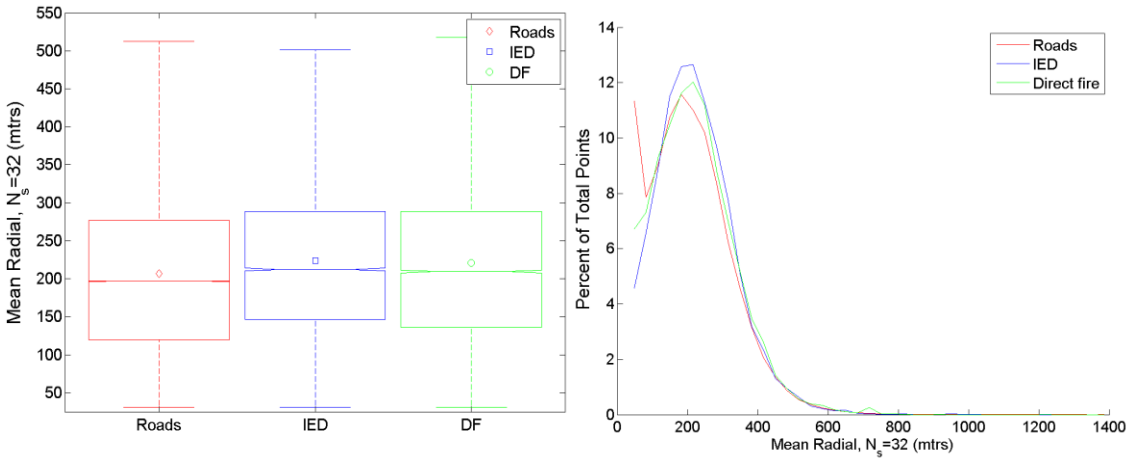



**Figure 114. Sparse viewshed mean radial ($N_s$ = 32).**

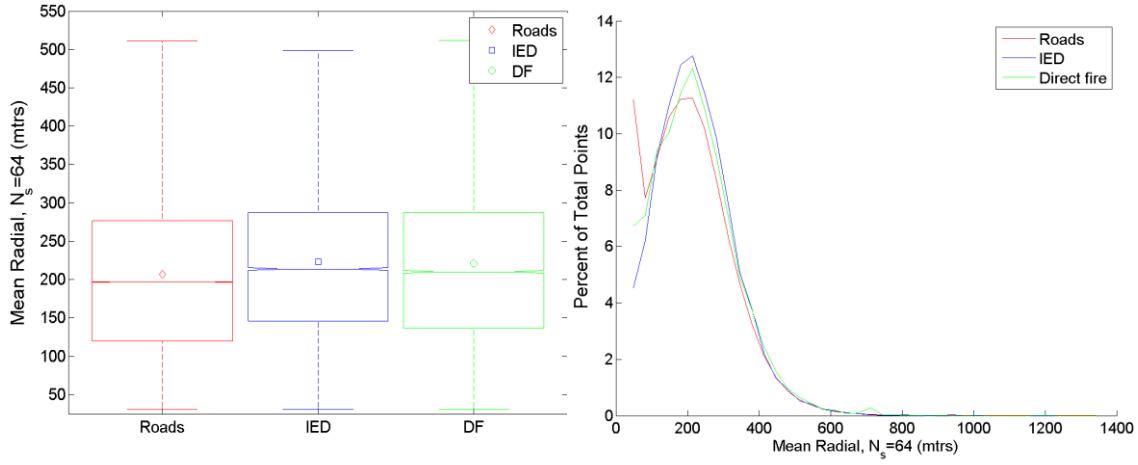



**Figure 115. Sparse viewshed mean radial ($N_s$ = 64).**

## Appendix B.13 Distance to Population Centers

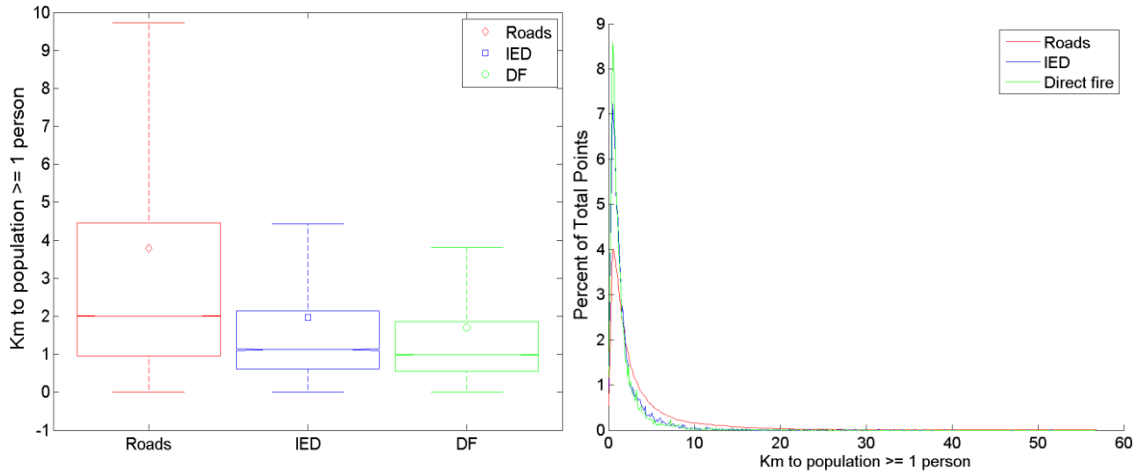



**Figure 116. Distance to nearest population center with more than 1 person.**

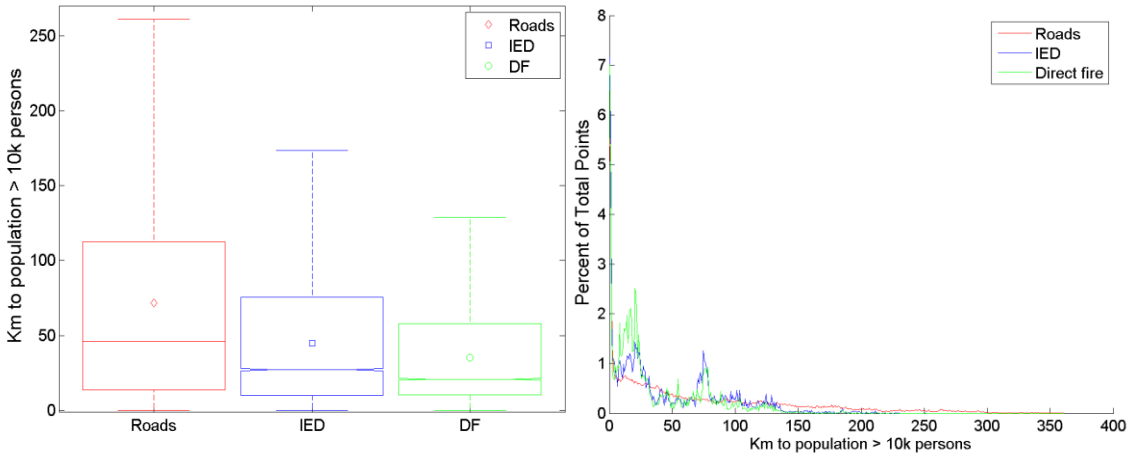



**Figure 117. Distance to nearest population center with more than 10,000 people.**

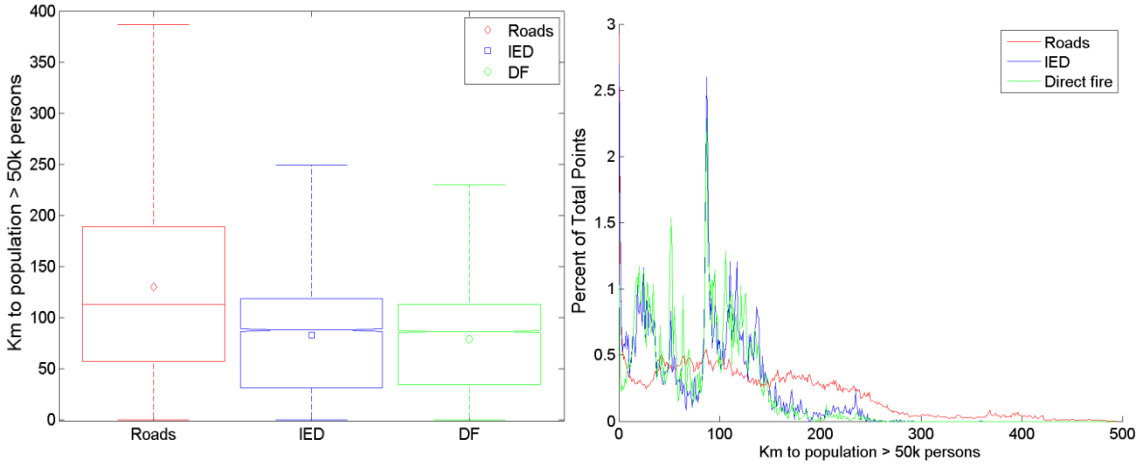



**Figure 118. Distance to nearest population center with more than 50,000 people.**

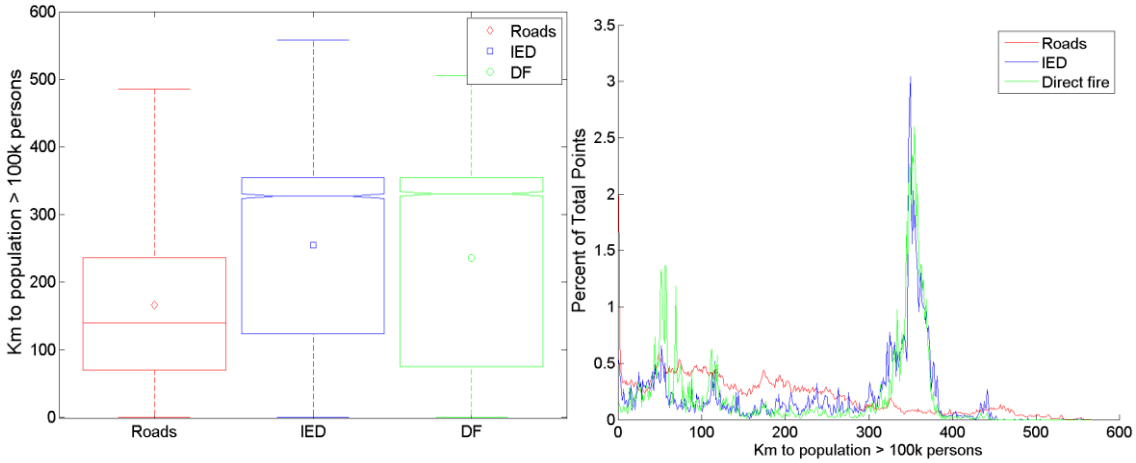



**Figure 119. Distance to nearest population center with more than 100,000 people.**

## Appendix B.14 Sparse Viewshed Planimetric Area

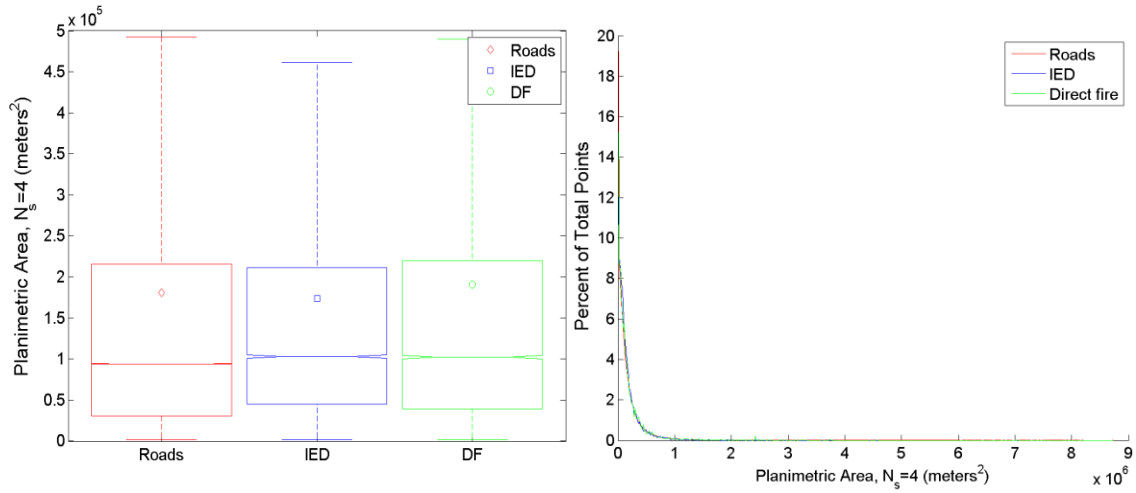



**Figure 120. sparse viewshed planimetric area ($N_s$ = 4).**

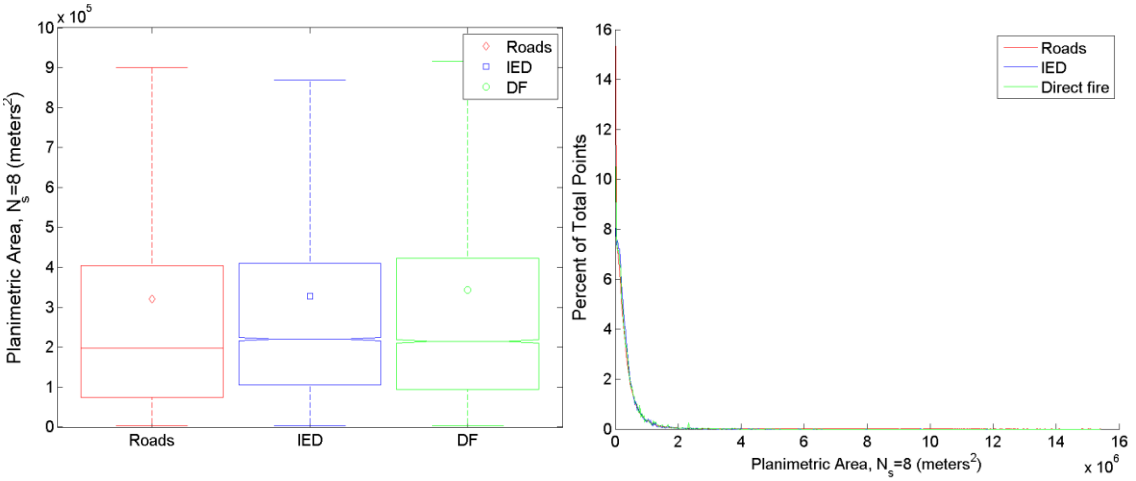



**Figure 121. Sparse viewshed planimetric area ($N_s$ = 8).**

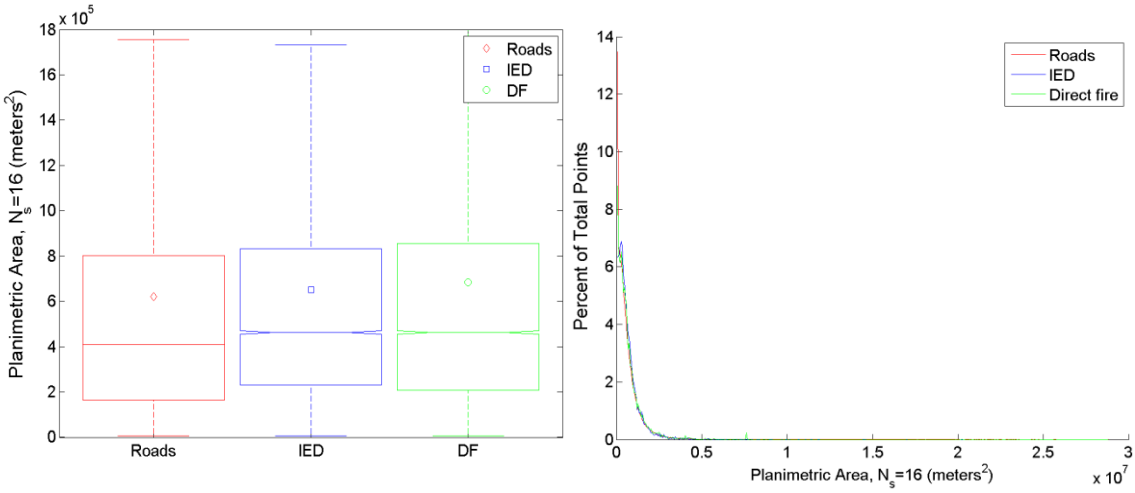



**Figure 122. Sparse viewshed planimetric area ($N_s$ = 16).**

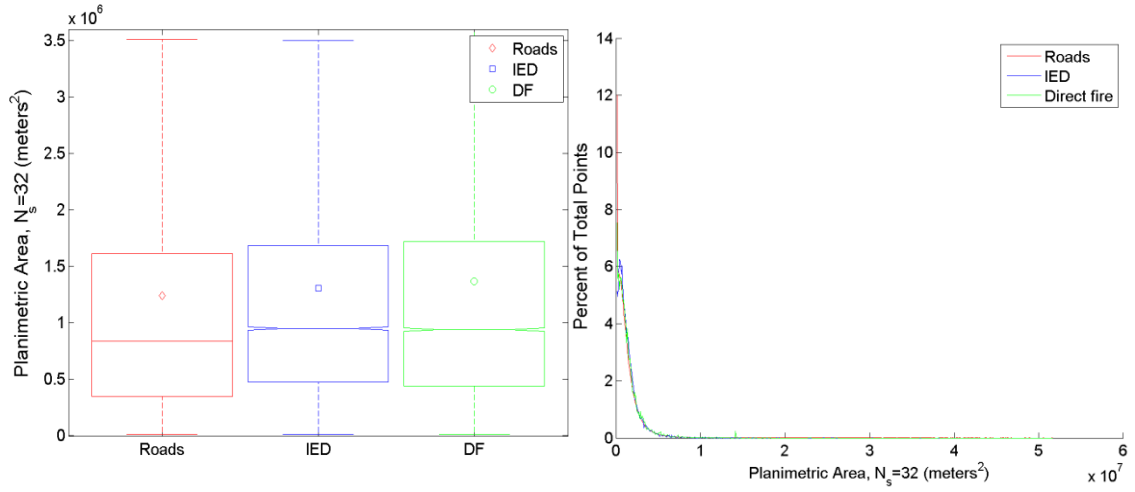



**Figure 123. Sparse viewshed planimetric area ($N_s$ = 32).**

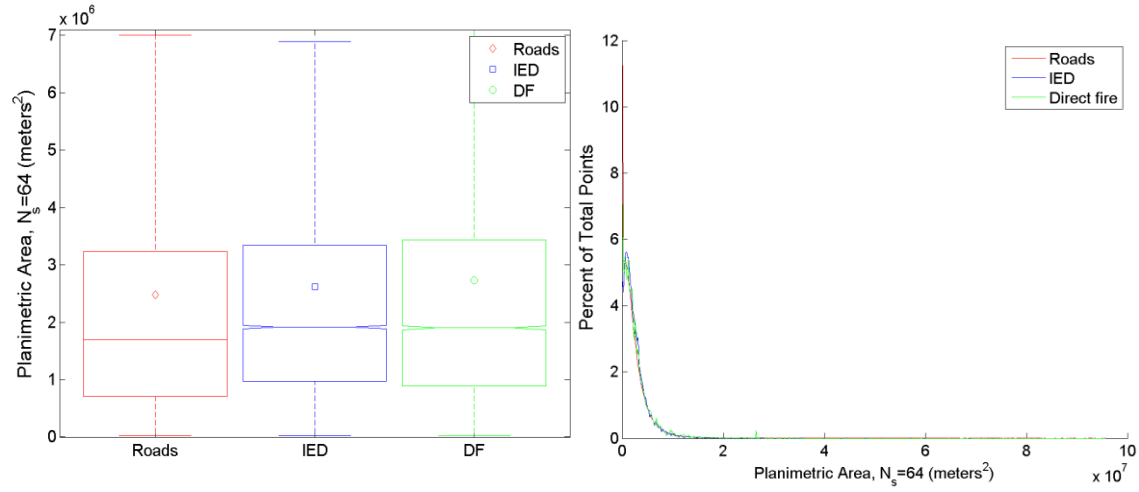



**Figure 124. Sparse viewshed planimetric area ($N_s$ = 64).**

## Appendix B.15 Sparse Viewshed Rugosity

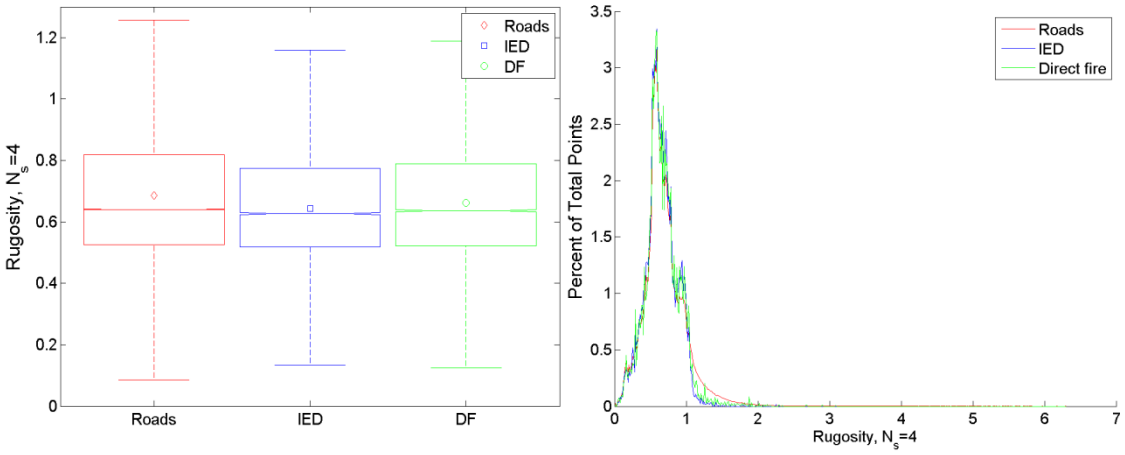



**Figure 125. Sparse viewshed rugosity ($N_s$ = 4).**

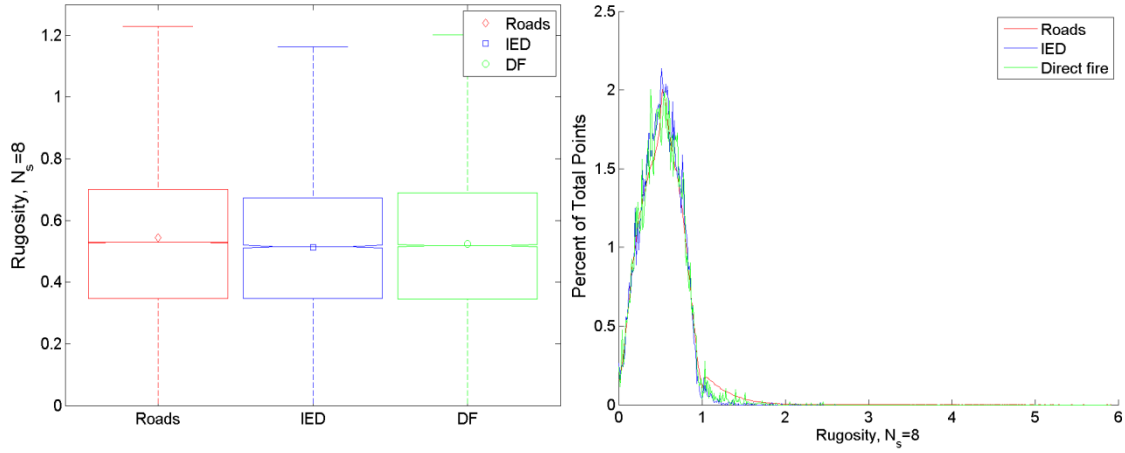



**Figure 126. Sparse viewshed rugosity ($N_s$ = 8).**

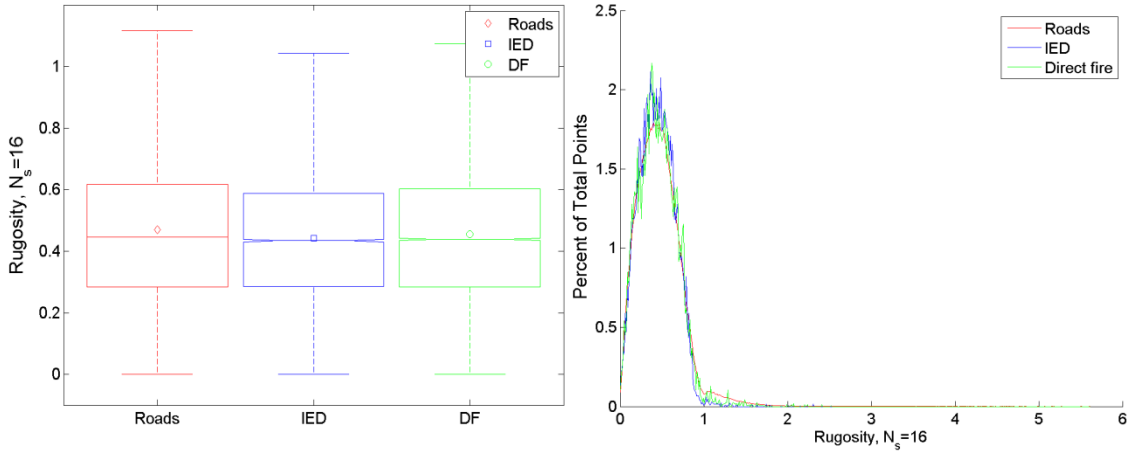



**Figure 127. Sparse viewshed rugosity ($N_s$ = 16).**

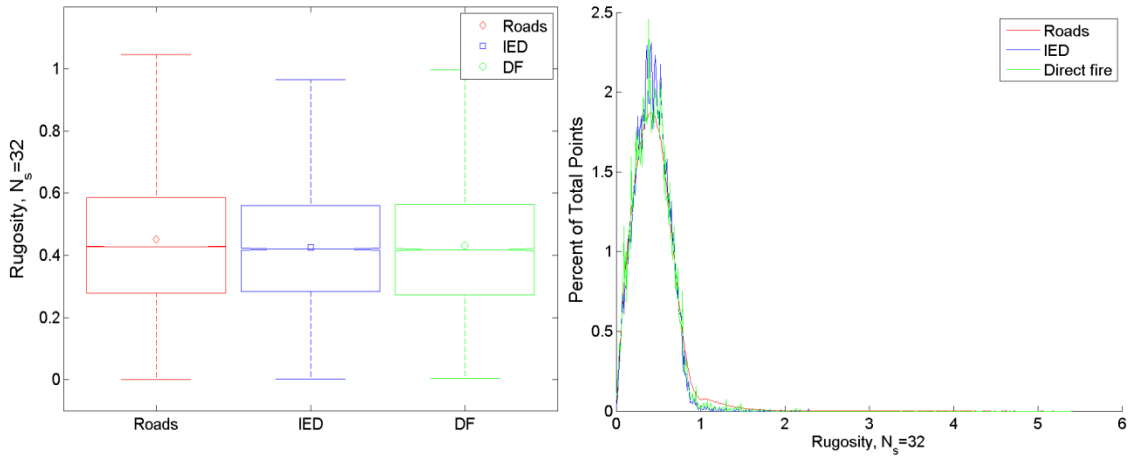



**Figure 128. Sparse viewshed rugosity ($N_s$ = 32).**

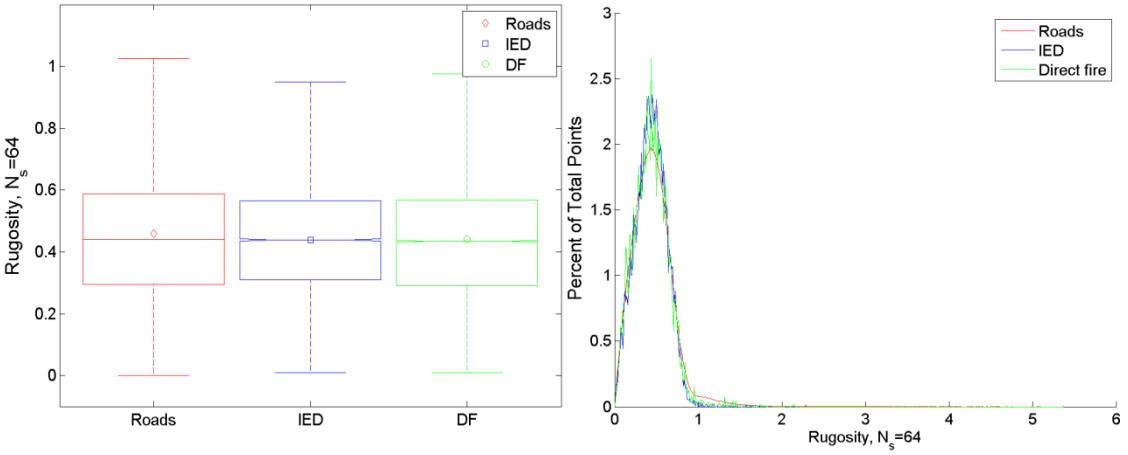



**Figure 129. Sparse viewshed rugosity ($N_s$ = 64).**

## Appendix B.16 Sparse Viewshed Shape Complexity Index

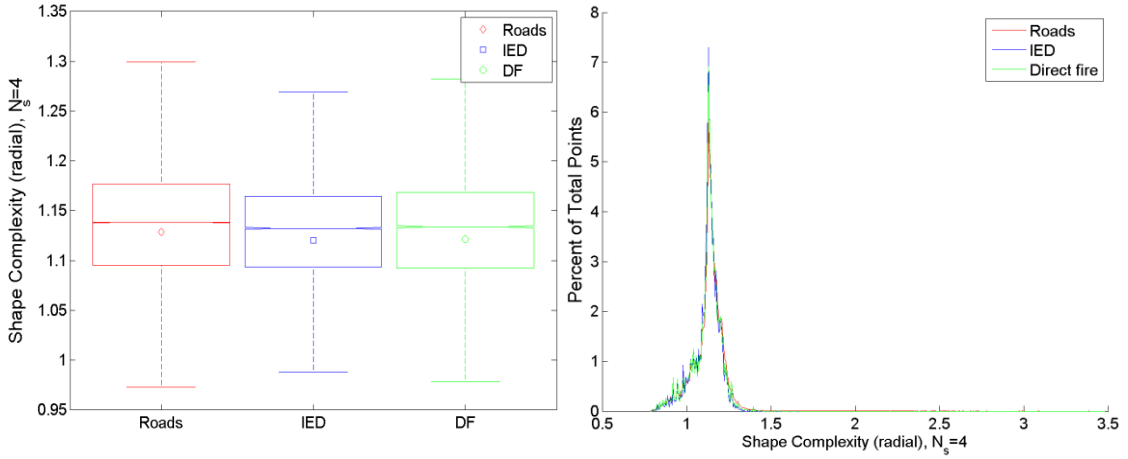



**Figure 130. Sparse viewshed shape complexity index ($N_s$ = 4).**

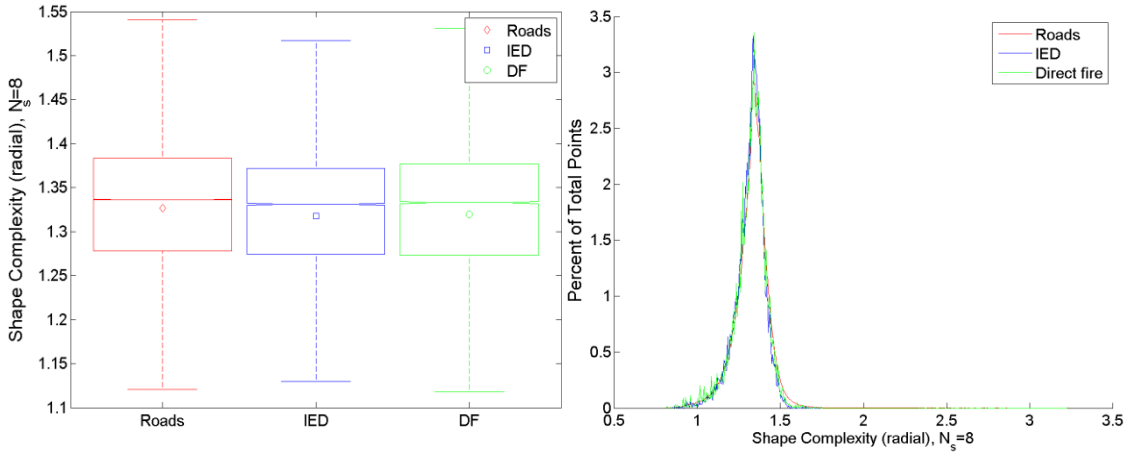



**Figure 131. Sparse viewshed shape complexity index ($N_s = 8$).**

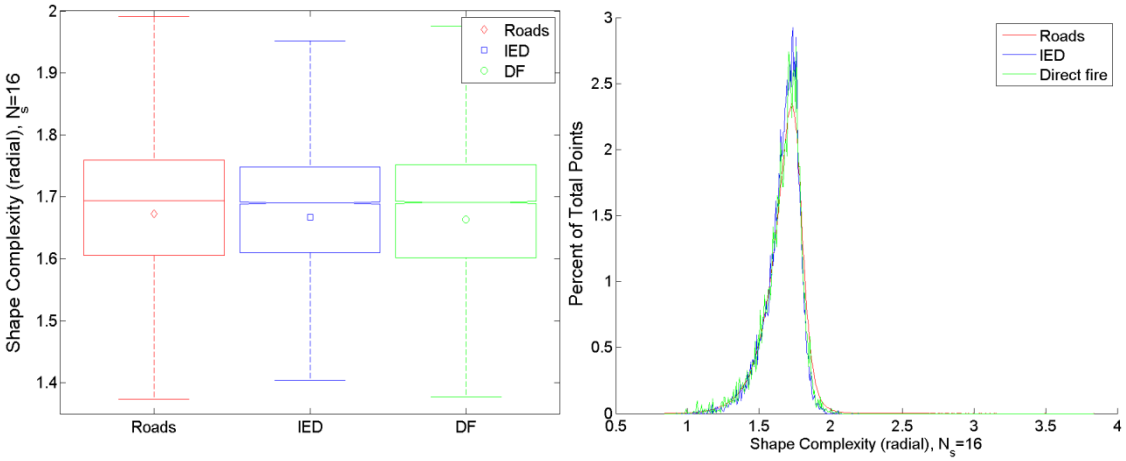



**Figure 132. Sparse viewshed shape complexity index ($N_s = 16$).**

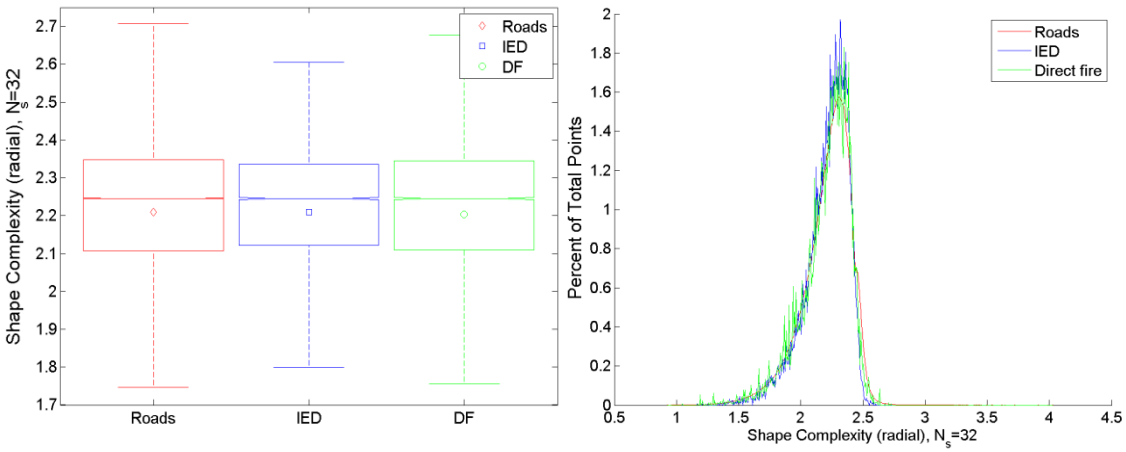



**Figure 133. Sparse viewshed shape complexity index ($N_s = 32$).**

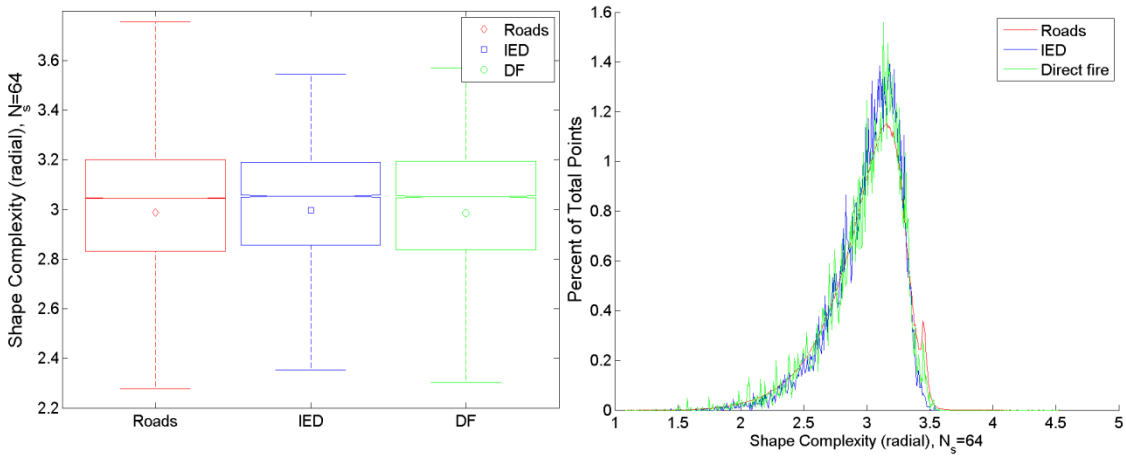



**Figure 134. Sparse viewshed shape complexity index ($N_s$ = 64).**

## APPENDIX C. SUMMARY OF FEATURES

The following tables summarize key statistics of the conflict event and road datasets.

**Table 12. Statistics for Road Points**

| Road Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Elevation | 1500.567 | 894.5486 | 0.543728 | 2.420101 | 5059 | 245 |
| Slope | 12.29649 | 11.22895 | 1.307544 | 4.18051 | 80.8767 | 0 |
| IW_convexity | 50.58728 | 3.467145 | 0.27602 | 7.149377 | 100 | 30.52326 |
| IW_texture | 171.7688 | 31.52016 | -0.47859 | 3.251345 | 268 | 0 |
| Elv_rng50 | 17.84263 | 17.63961 | 2.342669 | 11.46453 | 314 | 0 |
| Elv_rng100 | 30.36374 | 30.16361 | 2.123798 | 9.377894 | 486 | 0 |
| Elv_rng350 | 90.69964 | 90.76243 | 1.63301 | 5.771874 | 884 | 5 |
| Elv_rng500 | 120.7624 | 120.3921 | 1.524657 | 5.201717 | 1067 | 7 |
| Elv_rng1000 | 204.5278 | 202.0049 | 1.388893 | 4.580955 | 1480 | 10 |
| Rough_50 | 5.734752 | 5.809725 | 2.415465 | 12.15683 | 99.49794 | 0 |
| Rough_100 | 8.43356 | 8.84612 | 2.235514 | 10.32687 | 141.1226 | 0 |
| Rough_350 | 21.28197 | 23.7253 | 1.842704 | 7.022945 | 251.019 | 1.071471 |
| Rough_500 | 27.38109 | 30.47396 | 1.731273 | 6.292839 | 282.6358 | 1.182403 |
| Rough_1000 | 43.7797 | 48.01696 | 1.566073 | 5.328774 | 362.9134 | 1.575278 |
| Visidx100-350 | 173.6378 | 94.9768 | -0.13214 | 2.01938 | 360 | 0 |
| Visidx_350 | 200.8412 | 100.623 | -0.22429 | 2.080677 | 392 | 2 |
| Visidx_500 | 336.1799 | 187.3552 | 0.108157 | 2.140409 | 820 | 2 |
| Visidx_1000 | 820.063 | 557.0581 | 0.643432 | 2.798222 | 3172 | 2 |
| SCID100-350 | NaN | NaN | -0.75397 | 2.926884 | 5.352372 | 0.282095 |
| SCID_350 | 3.817166 | 1.188141 | -0.87124 | 3.083763 | 5.585192 | 0.398942 |
| SCID_500 | 4.902237 | 1.649369 | -0.59101 | 2.726625 | 8.077966 | 0.398942 |
| SCID_1000 | 7.522361 | 2.944933 | -0.13269 | 2.383551 | 15.88772 | 0.398942 |
| Short_rad_4 | 87.85215 | 68.96896 | 1.87485 | 8.553442 | 1081.062 | 30.88748 |
| Long_rad_4 | 418.7539 | 325.2703 | 1.950796 | 10.20155 | 2934.311 | 30.88748 |
| Mean_rad_4 | 222.8605 | 150.8685 | 1.300956 | 6.354747 | 1860.971 | 30.88748 |
| Local_op_4 | 0.145411 | 0.178123 | 3.137949 | 21.55073 | 4.192637 | 0 |
| Planimtrc_4 | 181005.4 | 299234.4 | 6.403355 | 78.60991 | 8741126 | 1908.073 |
| Rugosity_4 | 0.687045 | 0.291456 | 1.559916 | 11.25873 | 6.303492 | 0 |
| SCIF_4 | 1.128786 | 0.089746 | -0.1246 | 8.936473 | 3.486189 | 0.78235 |
| Short_rad_8 | 66.55558 | 46.47777 | 1.73964 | 7.199426 | 648.6371 | 30.88748 |
| Long_rad_8 | 492.7444 | 341.9749 | 1.766337 | 9.106218 | 2934.311 | 30.88748 |
| Mean_rad_8 | 211.2556 | 127.8879 | 1.020124 | 5.198145 | 1579.122 | 30.88748 |

| Road Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Local_op_8 | 0.147636 | 0.180521 | 3.202525 | 22.41808 | 4.528534 | 0.000458 |
| Planimtrc_8 | 321310.9 | 448361.5 | 5.669126 | 69.59668 | 15413320 | 3569.677 |
| Rugosity_8 | 0.545856 | 0.284494 | 1.308529 | 8.992094 | 5.920185 | 0 |
| SCIF_8 | 1.327031 | 0.099617 | -0.36451 | 6.428349 | 3.232216 | 0.804348 |
| Short_rad_16 | 56.8759 | 36.33826 | 1.705347 | 6.641047 | 494.1997 | 30.88748 |
| Long_rad_16 | 568.9883 | 367.5516 | 1.600696 | 8.075665 | 2934.311 | 30.88748 |
| Mean_rad_16 | 207.2524 | 119.6326 | 0.887791 | 4.712483 | 1498.043 | 30.88748 |
| Local_op_16 | 0.14788 | 0.181114 | 3.247771 | 22.90394 | 4.492111 | 0.001172 |
| Planimtrc_16 | 620615.5 | 795392.7 | 5.303083 | 63.93605 | 28862852 | 6769.165 |
| Rugosity_16 | 0.471404 | 0.265345 | 1.471321 | 9.785589 | 5.613903 | 0 |
| SCIF_16 | 1.672743 | 0.134303 | -0.84879 | 5.229183 | 3.840335 | 0.837967 |
| Short_rad_32 | 51.82218 | 30.52004 | 1.712476 | 6.511612 | 463.3122 | 30.88748 |
| Long_rad_32 | 646.2164 | 399.4636 | 1.455157 | 7.140774 | 2934.311 | 30.88748 |
| Mean_rad_32 | 206.5036 | 117.1211 | 0.833915 | 4.517314 | 1402.485 | 30.88748 |
| Local_op_32 | 0.148107 | 0.181391 | 3.257947 | 23.0757 | 4.51437 | 0.001517 |
| Planimtrc_32 | 1238309 | 1527504 | 5.130433 | 61.27438 | 51675859 | 13303.11 |
| Rugosity_32 | 0.451506 | 0.249126 | 1.556002 | 10.71974 | 5.414795 | 0 |
| SCIF_32 | 2.209487 | 0.198408 | -1.04606 | 4.735505 | 4.030713 | 0.919574 |
| Short_rad_64 | 49.37544 | 27.3026 | 1.670685 | 6.23432 | 339.7623 | 30.88748 |
| Long_rad_64 | 716.3153 | 432.0182 | 1.331562 | 6.378148 | 2934.311 | 30.88748 |
| Mean_rad_64 | 206.3157 | 116.2885 | 0.812963 | 4.443949 | 1358.566 | 30.88748 |
| Local_op_64 | 0.148148 | 0.181436 | 3.259188 | 23.07553 | 4.519934 | 0.001954 |
| Planimtrc_64 | 2473714 | 3004681 | 5.068907 | 60.3886 | 95729354 | 26481.33 |
| Rugosity_64 | 0.458616 | 0.241175 | 1.545595 | 10.99351 | 5.373364 | 0.000588 |
| SCIF_64 | 2.988666 | 0.29893 | -1.03653 | 4.406598 | 4.528361 | 1.07449 |
| Dist_pop_1 | 3866.659 | 5055.309 | 3.280635 | 18.8791 | 56877.3 | 0.799611 |
| Dist_pop_1k | 12018.52 | 22347.75 | 2.873837 | 11.86605 | 165470.1 | 0.799611 |
| Dist_pop_10k | 72976.86 | 73195.13 | 1.194508 | 3.810034 | 360925.8 | 0.799611 |
| Dist_pop_50k | 131591.1 | 94707.11 | 0.848839 | 3.589145 | 494683.7 | 0.799611 |
| Dist_pop100k | 165769.2 | 121818.8 | 0.882078 | 3.216328 | 566862.7 | 0.799611 |
| Min_CEA | 19.97321 | 39.10452 | 12.92167 | 393.3227 | 3394 | 1 |
| Max_CEA | 251.9178 | 416.94 | 6.272908 | 48.01058 | 4297 | 2 |
| Med_CEA | 83.38715 | 132.3518 | 6.841458 | 62.75816 | 3431 | 1 |
| RtVisMin_1k | 0.014877 | 0.021871 | 11.84692 | 200.5709 | 0.849057 | 0.000453 |
| RtVisMax_1k | 0.603916 | 0.184814 | -0.32498 | 2.982343 | 1 | 0.000582 |
| RtVisMed_1k | 0.182582 | 0.102707 | 0.729848 | 3.292173 | 0.88 | 0.000582 |
| RtVisMin_500 | 0.03024 | 0.03052 | 9.437664 | 136.5116 | 0.977273 | 0.001196 |
| RtVisMax500 | 0.726549 | 0.182017 | -0.84383 | 3.937267 | 1 | 0.001848 |

| Road Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| RtVisMed500 | 0.255945 | 0.133704 | 0.56446 | 2.837612 | 0.977273 | 0.001848 |
| RtVisMin_250 | 0.061097 | 0.041276 | 6.2992 | 72.69977 | 1 | 0.003472 |
| RtVisMax250 | 0.824756 | 0.170423 | -1.43312 | 5.704653 | 1 | 0.006098 |
| RtVisMed250 | 0.339879 | 0.161213 | 0.38499 | 2.535758 | 1 | 0.006098 |
| RtVisMin_100 | 0.15594 | 0.068637 | 2.642588 | 18.16312 | 1 | 0.014493 |
| RtVisMax100 | 0.924422 | 0.139425 | -2.53312 | 10.72765 | 1 | 0.026316 |
| RtVisMed100 | 0.480261 | 0.189432 | 0.076296 | 2.37194 | 1 | 0.020833 |

**Table 13. Statistics for IED Events**

| IED Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Elevation | 1189.741 | 516.1805 | 1.195342 | 3.562144 | 4231 | 284 |
| Slope | 5.960902 | 6.007035 | 2.954775 | 14.20707 | 57.00436 | 0 |
| IW_convexity | 50.50021 | 2.387105 | 0.214614 | 4.469146 | 66.27907 | 38.0814 |
| IW_texture | 185.0368 | 23.28743 | -0.40387 | 3.523254 | 261 | 69 |
| Elv_rng50 | 8.860173 | 7.52035 | 4.608615 | 37.40607 | 131 | 1 |
| Elv_rng100 | 13.89868 | 12.3034 | 4.693826 | 36.62869 | 204 | 2 |
| Elv_rng350 | 34.13291 | 36.99049 | 4.682419 | 34.68556 | 547 | 7 |
| Elv_rng500 | 44.05115 | 49.79456 | 4.384505 | 30.32021 | 663 | 9 |
| Elv_rng1000 | 73.97255 | 86.47978 | 3.587551 | 21.02291 | 1056 | 12 |
| Rough_50 | 2.810664 | 2.463042 | 4.6916 | 37.65946 | 41.45635 | 0.426401 |
| Rough_100 | 3.711215 | 3.581271 | 4.951779 | 40.19897 | 63.20649 | 0.700662 |
| Rough_350 | 7.02624 | 9.215696 | 5.019333 | 38.89898 | 154.6467 | 1.366325 |
| Rough_500 | 8.655027 | 12.0059 | 4.845296 | 36.23133 | 187.591 | 1.520224 |
| Rough_1000 | 13.58592 | 19.17193 | 4.150777 | 27.41496 | 245.6013 | 1.839131 |
| Visidx100-350 | 185.6878 | 87.87364 | -0.18106 | 2.140015 | 360 | 0 |
| Visidx_350 | 214.4752 | 91.72991 | -0.26132 | 2.227028 | 392 | 4 |
| Visidx_500 | 351.4775 | 180.5928 | 0.120343 | 2.1702 | 806 | 4 |
| Visidx_1000 | 811.8062 | 567.0791 | 0.73444 | 2.8867 | 2943 | 4 |
| SCID100-350 | NaN | NaN | -0.77079 | 3.112951 | 5.352372 | 0.282095 |
| SCID_350 | 4.006368 | 1.008209 | -0.8829 | 3.44103 | 5.585192 | 0.56419 |
| SCID_500 | 5.069513 | 1.506619 | -0.51472 | 2.722967 | 8.00871 | 0.56419 |
| SCID_1000 | 7.480773 | 2.939417 | 0.00766 | 2.292969 | 15.30348 | 0.56419 |
| Short_rad_4 | 93.87285 | 70.00596 | 1.909031 | 9.484288 | 957.5119 | 30.88748 |
| Long_rad_4 | 424.1586 | 280.8768 | 1.994782 | 11.18061 | 2934.311 | 30.88748 |
| Mean_rad_4 | 232.2845 | 134.2474 | 1.205886 | 5.800196 | 1258.665 | 30.88748 |
| Local_op_4 | 0.059758 | 0.080886 | 5.322349 | 52.79569 | 1.359774 | 0 |

| IED Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Planimtrc_4 | 174157.1 | 244022.5 | 5.429662 | 55.67555 | 4540737 | 1908.073 |
| Rugosity_4 | 0.642722 | 0.216575 | 0.354551 | 4.804443 | 2.282296 | 0 |
| SCIF_4 | 1.120155 | 0.077516 | -1.01414 | 4.691009 | 1.410832 | 0.798868 |
| Short_rad_8 | 72.11338 | 48.61465 | 1.698557 | 7.617441 | 555.9746 | 30.88748 |
| Long_rad_8 | 506.9798 | 299.1997 | 1.835823 | 10.03663 | 2934.311 | 30.88748 |
| Mean_rad_8 | 226.2865 | 117.4495 | 0.991791 | 5.003164 | 1077.201 | 30.88748 |
| Local_op_8 | 0.060769 | 0.082075 | 5.63435 | 61.95781 | 1.663295 | 0.001457 |
| Planimtrc_8 | 328125.1 | 389897.4 | 4.877909 | 56.08882 | 9702172 | 3569.677 |
| Rugosity_8 | 0.512435 | 0.227316 | 0.309122 | 4.177166 | 2.425573 | 0 |
| SCIF_8 | 1.317845 | 0.086136 | -0.92441 | 5.105927 | 1.653756 | 0.852437 |
| Short_rad_16 | 61.38465 | 38.18321 | 1.594601 | 6.427533 | 339.7623 | 30.88748 |
| Long_rad_16 | 590.6188 | 322.0858 | 1.664434 | 8.798545 | 2934.311 | 30.88748 |
| Mean_rad_16 | 224.1002 | 110.6749 | 0.908346 | 4.969063 | 1104.227 | 30.88748 |
| Local_op_16 | 0.061132 | 0.082036 | 5.615595 | 61.46527 | 1.675436 | 0.002342 |
| Planimtrc_16 | 651005.1 | 727405.5 | 4.99593 | 55.56582 | 14577227 | 6769.165 |
| Rugosity_16 | 0.441361 | 0.209275 | 0.46621 | 4.412503 | 2.308768 | 0 |
| SCIF_16 | 1.667039 | 0.118685 | -1.1348 | 5.096488 | 2.062202 | 1.042637 |
| Short_rad_32 | 55.92736 | 32.65441 | 1.610797 | 6.399642 | 277.9873 | 30.88748 |
| Long_rad_32 | 670.5452 | 348.6348 | 1.519722 | 7.758947 | 2934.311 | 30.88748 |
| Mean_rad_32 | 223.7752 | 108.2676 | 0.856616 | 4.84667 | 1008.669 | 30.88748 |
| Local_op_32 | 0.061157 | 0.082141 | 5.712708 | 64.01138 | 1.691624 | 0.002632 |
| Planimtrc_32 | 1307310 | 1412254 | 5.038128 | 58.74511 | 30304717 | 13303.11 |
| Rugosity_32 | 0.4264 | 0.195578 | 0.530602 | 5.117283 | 2.28104 | 0.002348 |
| SCIF_32 | 2.208976 | 0.176938 | -1.21226 | 4.989893 | 2.605695 | 1.162805 |
| Short_rad_64 | 53.34624 | 29.65709 | 1.571263 | 6.196626 | 277.9873 | 30.88748 |
| Long_rad_64 | 744.8215 | 383.6695 | 1.462154 | 7.206687 | 2934.311 | 30.88748 |
| Mean_rad_64 | 223.5386 | 107.386 | 0.831561 | 4.755836 | 1006.256 | 30.88748 |
| Local_op_64 | 0.061176 | 0.082221 | 5.720145 | 64.28753 | 1.711353 | 0.002826 |
| Planimtrc_64 | 2616312 | 2791769 | 4.85154 | 53.25151 | 57114787 | 26481.33 |
| Rugosity_64 | 0.438592 | 0.18791 | 0.462449 | 5.409028 | 2.278868 | 0.009078 |
| SCIF_64 | 2.997456 | 0.264928 | -1.15723 | 4.735597 | 3.545446 | 1.447101 |
| Dist_pop_1 | 1970.09 | 3025.814 | 6.955584 | 82.21161 | 62357.3 | 8.145166 |
| Dist_pop_1k | 4775.628 | 12941.26 | 5.415877 | 41.72045 | 192750.7 | 8.145166 |
| Dist_pop_10k | 44790.97 | 42667.25 | 1.00429 | 3.476587 | 230029.5 | 8.145166 |
| Dist_pop_50k | 82942.68 | 55757.04 | 0.484679 | 2.966735 | 360583 | 38.99919 |
| Dist_pop100k | 255149.6 | 129590.4 | -0.71614 | 1.960563 | 557783.2 | 139.1335 |
| Min_CEA | 15.98684 | 28.32625 | 7.408402 | 102.6326 | 769 | 1 |
| Max_CEA | 210.9615 | 246.4451 | 5.496054 | 49.01561 | 4297 | 1 |

| IED Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Med_CEA | 65.95197 | 76.60394 | 4.774067 | 40.36239 | 1357 | 1 |
| RtVisMin_1k | 0.126199 | 0.261115 | 2.616346 | 8.707304 | 1 | 0.000518 |
| RtVisMax_1k | 0.655776 | 0.216014 | -0.06175 | 2.469475 | 1 | 0.004367 |
| RtVisMed_1k | 0.251898 | 0.245152 | 2.030872 | 6.560558 | 1 | 0.003802 |
| RtVisMin_500 | 0.236852 | 0.340464 | 1.509931 | 3.722207 | 1 | 0.001678 |
| RtVisMax500 | 0.794972 | 0.195926 | -0.81358 | 3.290479 | 1 | 0.007692 |
| RtVisMed500 | 0.379401 | 0.300473 | 1.125179 | 3.029845 | 1 | 0.006289 |
| RtVisMin_250 | 0.38808 | 0.388385 | 0.718764 | 1.822657 | 1 | 0.006803 |
| RtVisMax250 | 0.894392 | 0.159392 | -1.80274 | 6.542154 | 1 | 0.032258 |
| RtVisMed250 | 0.52769 | 0.326781 | 0.409904 | 1.690512 | 1 | 0.016667 |
| RtVisMin_100 | 0.696318 | 0.371742 | -0.55682 | 1.525536 | 1 | 0.023256 |
| RtVisMax100 | 0.974347 | 0.089017 | -4.57539 | 28.23482 | 1 | 0.083333 |
| RtVisMed100 | 0.792028 | 0.283912 | -0.90026 | 2.263288 | 1 | 0.033333 |

**Table 14. Statistics for Direct Fire Events**

| DF Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Elevation | 1239.97 | 557.3591 | 1.172477 | 3.269075 | 4151 | 275 |
| Slope | 8.178474 | 9.209225 | 2.202922 | 7.689192 | 64.3739 | 0 |
| IW_convexity | 50.78444 | 2.767828 | 0.341509 | 4.972919 | 69.18605 | 35.75581 |
| IW_texture | 179.9506 | 28.21021 | -0.62173 | 4.025256 | 255 | 47 |
| Elv_rng50 | 11.94124 | 13.05931 | 2.970686 | 13.34795 | 116 | 1 |
| Elv_rng100 | 19.69348 | 22.53343 | 2.863105 | 12.18684 | 188 | 2 |
| Elv_rng350 | 56.02511 | 76.16447 | 2.742375 | 11.17331 | 571 | 7 |
| Elv_rng500 | 74.25974 | 103.065 | 2.605197 | 10.1242 | 769 | 8 |
| Elv_rng1000 | 126.0677 | 171.9339 | 2.249472 | 7.89859 | 1351 | 12 |
| Rough_50 | 3.811751 | 4.2688 | 3.014123 | 13.6756 | 40.07682 | 0.389249 |
| Rough_100 | 5.365653 | 6.525197 | 2.941597 | 12.74078 | 55.62539 | 0.552669 |
| Rough_350 | 12.44763 | 19.00739 | 2.874951 | 12.27379 | 158.8052 | 1.306791 |
| Rough_500 | 15.99117 | 25.3049 | 2.824251 | 11.9456 | 218.367 | 1.480848 |
| Rough_1000 | 25.8792 | 40.7347 | 2.487302 | 9.286029 | 337.7308 | 1.844114 |
| Visidx100-350 | 180.6762 | 91.50847 | -0.18309 | 2.100827 | 360 | 0 |
| Visidx_350 | 208.9338 | 96.00504 | -0.26845 | 2.184552 | 392 | 4 |
| Visidx_500 | 347.2455 | 185.3923 | 0.08659 | 2.167492 | 816 | 4 |
| Visidx_1000 | 845.3988 | 585.3677 | 0.632782 | 2.67354 | 3152 | 4 |
| SCID100-350 | NaN | NaN | -0.7774 | 3.048465 | 5.352372 | 0.282095 |
| SCID_350 | 3.927529 | 1.095906 | -0.93006 | 3.422014 | 5.585192 | 0.56419 |

| DF Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| SCID_500 | 5.008353 | 1.596708 | -0.61598 | 2.859334 | 8.058239 | 0.56419 |
| SCID_1000 | 7.61272 | 3.053154 | -0.12254 | 2.32551 | 15.83756 | 0.56419 |
| Short_rad_4 | 91.54477 | 70.65889 | 1.916574 | 8.422847 | 679.5246 | 30.88748 |
| Long_rad_4 | 432.9138 | 323.5822 | 2.21322 | 11.68228 | 2934.311 | 30.88748 |
| Mean_rad_4 | 232.4906 | 147.4341 | 1.328615 | 6.070495 | 1413.102 | 30.88748 |
| Local_op_4 | 0.084963 | 0.12842 | 4.37544 | 32.41407 | 1.853502 | 0.000506 |
| Planimtrc_4 | 191101.9 | 311498.5 | 5.611846 | 54.29834 | 7054350 | 1908.073 |
| Rugosity_4 | 0.661678 | 0.246279 | 1.091447 | 8.373619 | 3.08116 | 0 |
| SCIF_4 | 1.12125 | 0.084467 | -0.78846 | 4.821488 | 1.659523 | 0.793605 |
| Short_rad_8 | 70.05427 | 48.26989 | 1.768018 | 7.634148 | 494.1997 | 30.88748 |
| Long_rad_8 | 515.6846 | 339.143 | 1.95246 | 9.925631 | 2934.311 | 30.88748 |
| Mean_rad_8 | 223.782 | 124.5285 | 0.973981 | 4.845281 | 1343.605 | 30.88748 |
| Local_op_8 | 0.086389 | 0.131173 | 4.532036 | 34.4163 | 1.82922 | 0.001515 |
| Planimtrc_8 | 342874.2 | 443987.1 | 5.006058 | 59.0276 | 11788326 | 3569.677 |
| Rugosity_8 | 0.524474 | 0.254098 | 0.883732 | 6.585551 | 2.599429 | 0 |
| SCIF_8 | 1.319373 | 0.095203 | -0.93913 | 5.47292 | 1.738768 | 0.818671 |
| Short_rad_16 | 59.66794 | 37.1889 | 1.564908 | 6.062438 | 339.7623 | 30.88748 |
| Long_rad_16 | 601.1742 | 370.5594 | 1.822419 | 9.073959 | 2934.311 | 30.88748 |
| Mean_rad_16 | 221.924 | 118.4034 | 0.940168 | 4.981722 | 1270.248 | 30.88748 |
| Local_op_16 | 0.086577 | 0.131033 | 4.520773 | 34.05139 | 1.865643 | 0.002342 |
| Planimtrc_16 | 685665.5 | 852809.9 | 4.978382 | 52.65761 | 20257471 | 6769.165 |
| Rugosity_16 | 0.45431 | 0.237426 | 1.042668 | 7.099692 | 2.523818 | 0 |
| SCIF_16 | 1.663809 | 0.132212 | -1.23675 | 5.690666 | 2.191069 | 0.926427 |
| Short_rad_32 | 54.86252 | 32.02801 | 1.574108 | 6.080916 | 339.7623 | 30.88748 |
| Long_rad_32 | 682.435 | 402.3281 | 1.695075 | 8.127751 | 2934.311 | 30.88748 |
| Mean_rad_32 | 221.0894 | 115.7138 | 0.897515 | 4.936341 | 1230.673 | 30.88748 |
| Local_op_32 | 0.086496 | 0.130627 | 4.503245 | 33.54321 | 1.795833 | 0.002536 |
| Planimtrc_32 | 1367396 | 1639130 | 4.997501 | 55.40848 | 38216401 | 13303.11 |
| Rugosity_32 | 0.431424 | 0.220303 | 1.105502 | 7.612239 | 2.183659 | 0.003546 |
| SCIF_32 | 2.203323 | 0.197384 | -1.22504 | 5.221941 | 2.72501 | 1.044676 |
| Short_rad_64 | 52.61044 | 29.49649 | 1.55894 | 6.011126 | 339.7623 | 30.88748 |
| Long_rad_64 | 759.4668 | 438.0354 | 1.558668 | 7.210818 | 2934.311 | 30.88748 |
| Mean_rad_64 | 220.8066 | 115.0348 | 0.875594 | 4.827426 | 1207.99 | 30.88748 |
| Local_op_64 | 0.086506 | 0.130681 | 4.512222 | 33.68811 | 1.819103 | 0.002668 |
| Planimtrc_64 | 2730956 | 3228337 | 4.916075 | 54.12197 | 73489390 | 26481.33 |
| Rugosity_64 | 0.4401 | 0.212687 | 1.071789 | 7.815886 | 2.198489 | 0.009 |
| SCIF_64 | 2.985627 | 0.293502 | -1.19028 | 4.828701 | 3.570024 | 1.327981 |
| Dist_pop_1 | 1700.359 | 2487.212 | 6.552423 | 79.11162 | 60470.71 | 6.711227 |

| DF Features | Mean | Std Dev. | Skewness | Kurtosis | Max | Min |
|---|---|---|---|---|---|---|
| Dist_pop_1k | 2617.723 | 7444.843 | 8.182352 | 87.77123 | 148677.1 | 6.711227 |
| Dist_pop_10k | 35394.62 | 35869.29 | 1.364658 | 4.598799 | 250508.8 | 20.60933 |
| Dist_pop_50k | 79040.08 | 47634.61 | 0.439541 | 3.084528 | 290180.5 | 45.84043 |
| Dist_pop100k | 235779.4 | 137015.1 | -0.41086 | 1.371092 | 504834.1 | 214.2961 |
| Min_CEA | 15.15311 | 21.41143 | 4.59485 | 40.23002 | 367 | 1 |
| Max_CEA | 192.7419 | 168.7335 | 5.057386 | 57.53948 | 3849 | 1 |
| Med_CEA | 59.22232 | 55.58704 | 3.862144 | 40.28346 | 1166 | 1 |
| RtVisMin_1k | 0.141893 | 0.260763 | 2.488321 | 8.165724 | 1 | 0.000726 |
| RtVisMax_1k | 0.674881 | 0.214215 | -0.22523 | 2.539017 | 1 | 0.007463 |
| RtVisMed_1k | 0.264446 | 0.247878 | 1.862538 | 5.937132 | 1 | 0.003846 |
| RtVisMin_500 | 0.277799 | 0.340138 | 1.282869 | 3.182472 | 1 | 0.002632 |
| RtVisMax500 | 0.818447 | 0.189101 | -1.05744 | 3.910993 | 1 | 0.018519 |
| RtVisMed500 | 0.41733 | 0.304879 | 0.860691 | 2.505724 | 1 | 0.011494 |
| RtVisMin_250 | 0.455908 | 0.379136 | 0.4267 | 1.576591 | 1 | 0.00885 |
| RtVisMax250 | 0.919258 | 0.141838 | -2.28398 | 9.050088 | 1 | 0.045455 |
| RtVisMed250 | 0.589679 | 0.319247 | 0.081273 | 1.573079 | 1 | 0.015152 |
| RtVisMin_100 | 0.750182 | 0.339886 | -0.87639 | 2.106256 | 1 | 0.015625 |
| RtVisMax100 | 0.983039 | 0.07348 | -6.07345 | 48.09473 | 1 | 0.058824 |
| RtVisMed100 | 0.838916 | 0.254278 | -1.33644 | 3.416514 | 1 | 0.02381 |

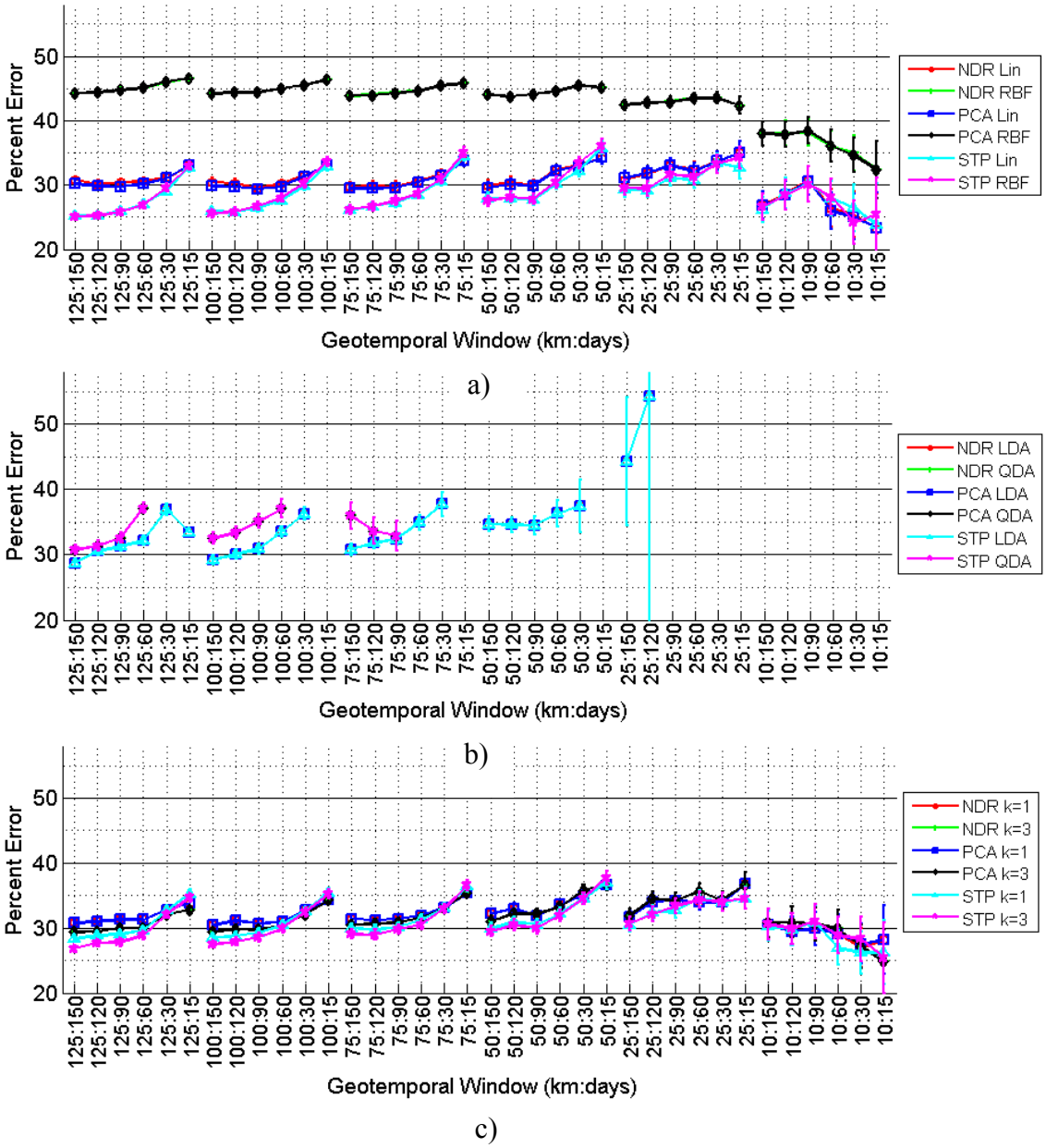



**Figure 135. The impact of combined geographic and temporal windows on classification accuracy of DF events using default parameters; a) DF SVM; b) DF DA; c) DF kNN.**

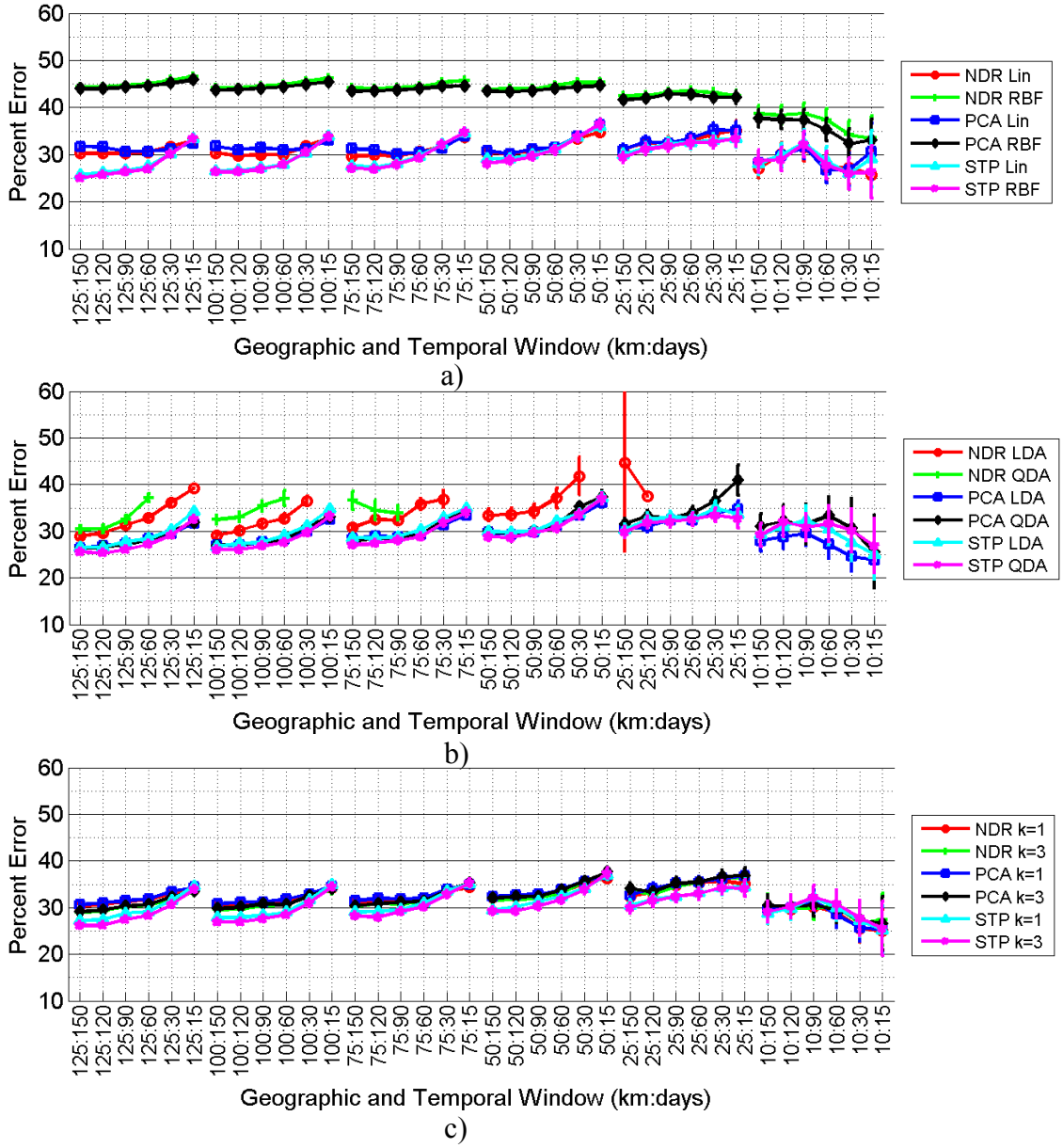



**Figure 136. The impact of combined geographic and temporal windows on classification accuracy of DF events when the cumulative variance of the PCA components ≤ 95% and the maximum p-value of the stepwise-selected features ≤ 0.05; a) DF SVM; b) DF DA; c) DF kNN.**

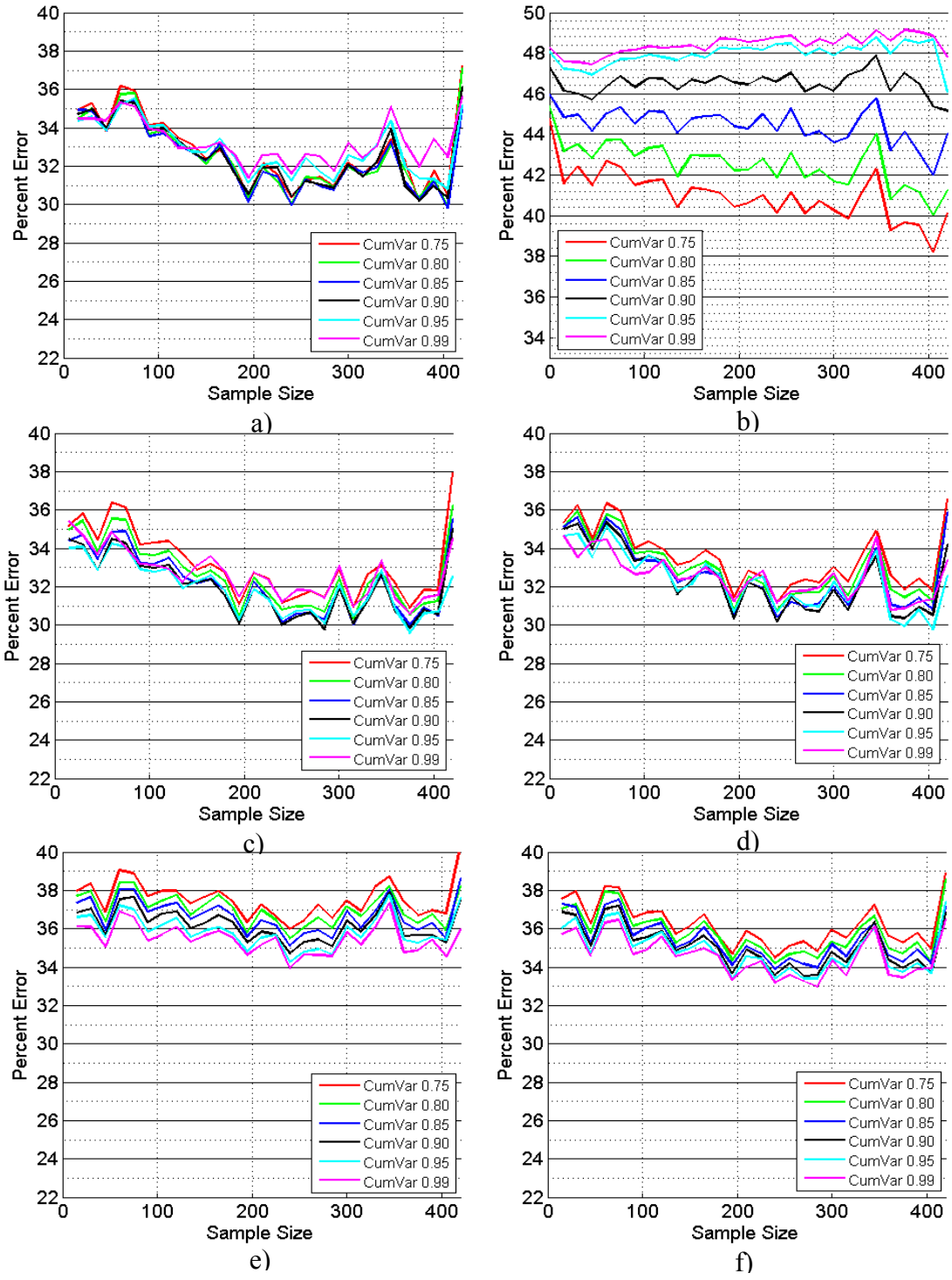



**Figure 137. The impact of varying cumulative variance and *penter* parameters on the classification of IED data; a) PCA with SVM linear kernel; b) PCA with SVM RBF kernel; c) PCA with LDA; d) PCA with QDA; e) PCA with kNN (k=1); f) PCA with kNN (k=3).**

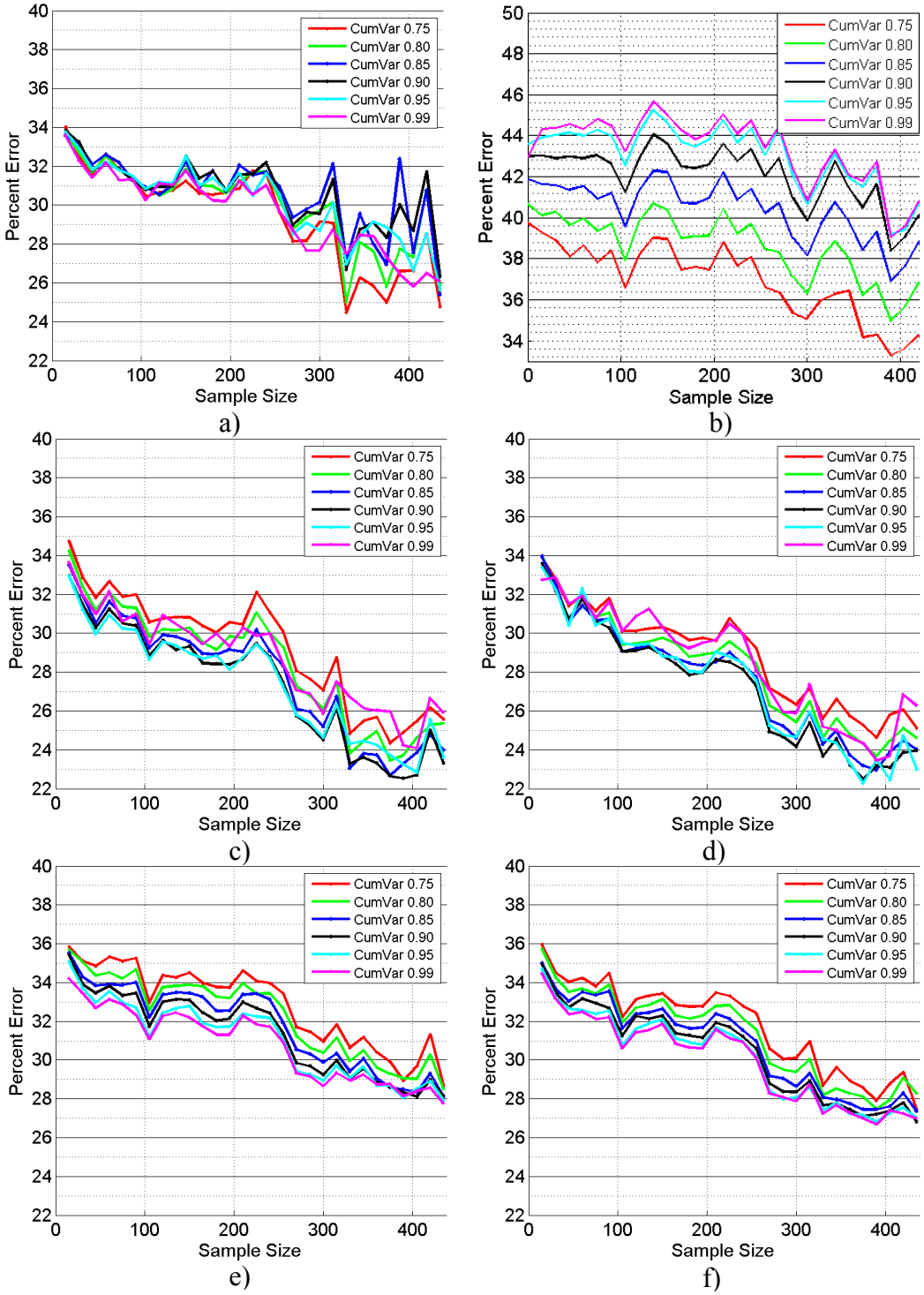



**Figure 138. The impact of varying cumulative variance and *penter* parameters on the classification of DF data; a) PCA with SVM linear kernel; b) PCA with SVM RBF kernel; c) PCA with LDA; d) PCA with QDA; e) PCA with kNN (k=1); f) PCA with kNN (k=3).**

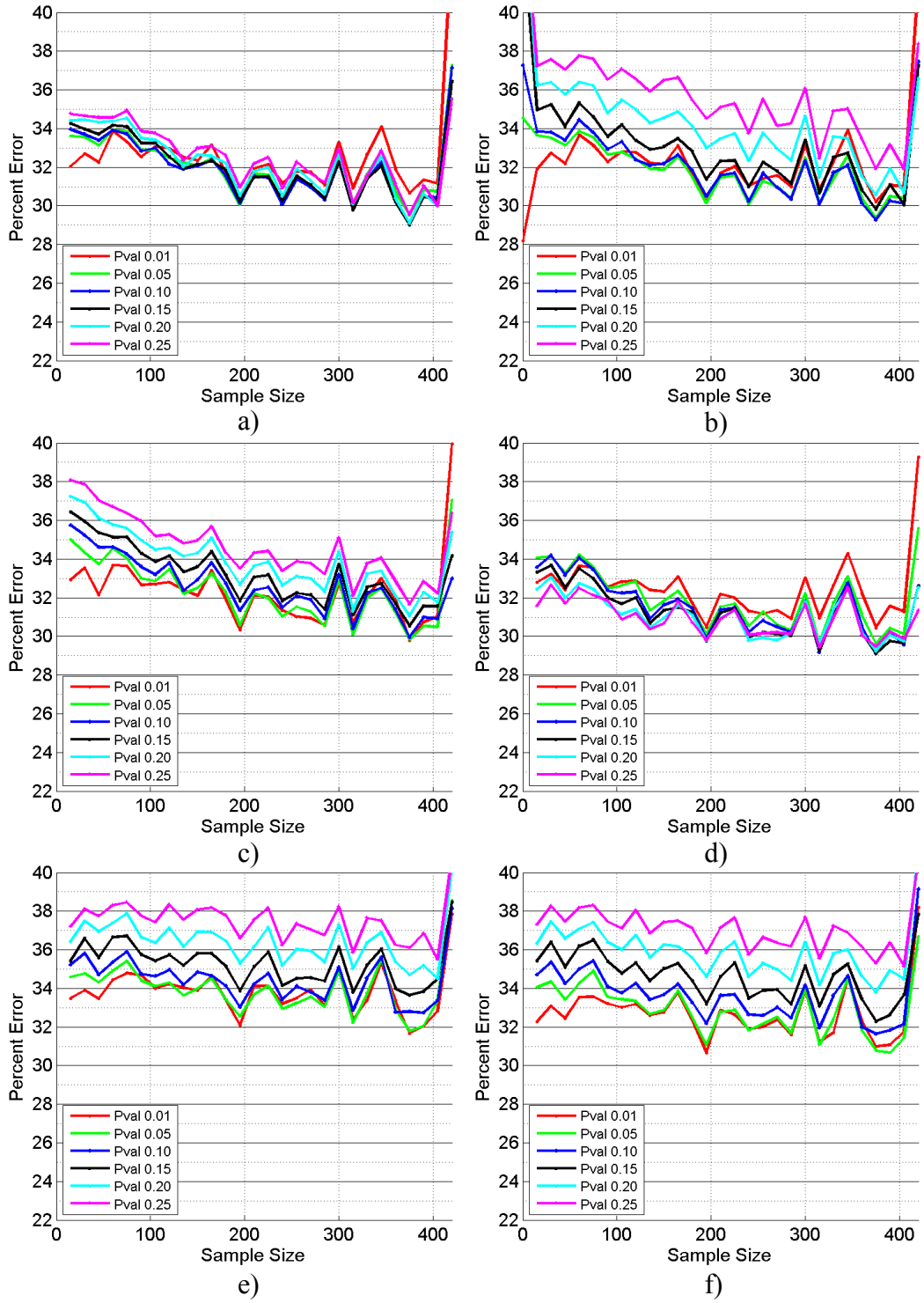



**Figure 139. The impact of varying cumulative variance and *penter* parameters on the classification of IED data; a) STP with SVM linear kernel; b) STP with SVM RBF kernel; c) STP with LDA; d) STP with QDA; e) STP with kNN (k=1); f) STP with kNN (k=3).**

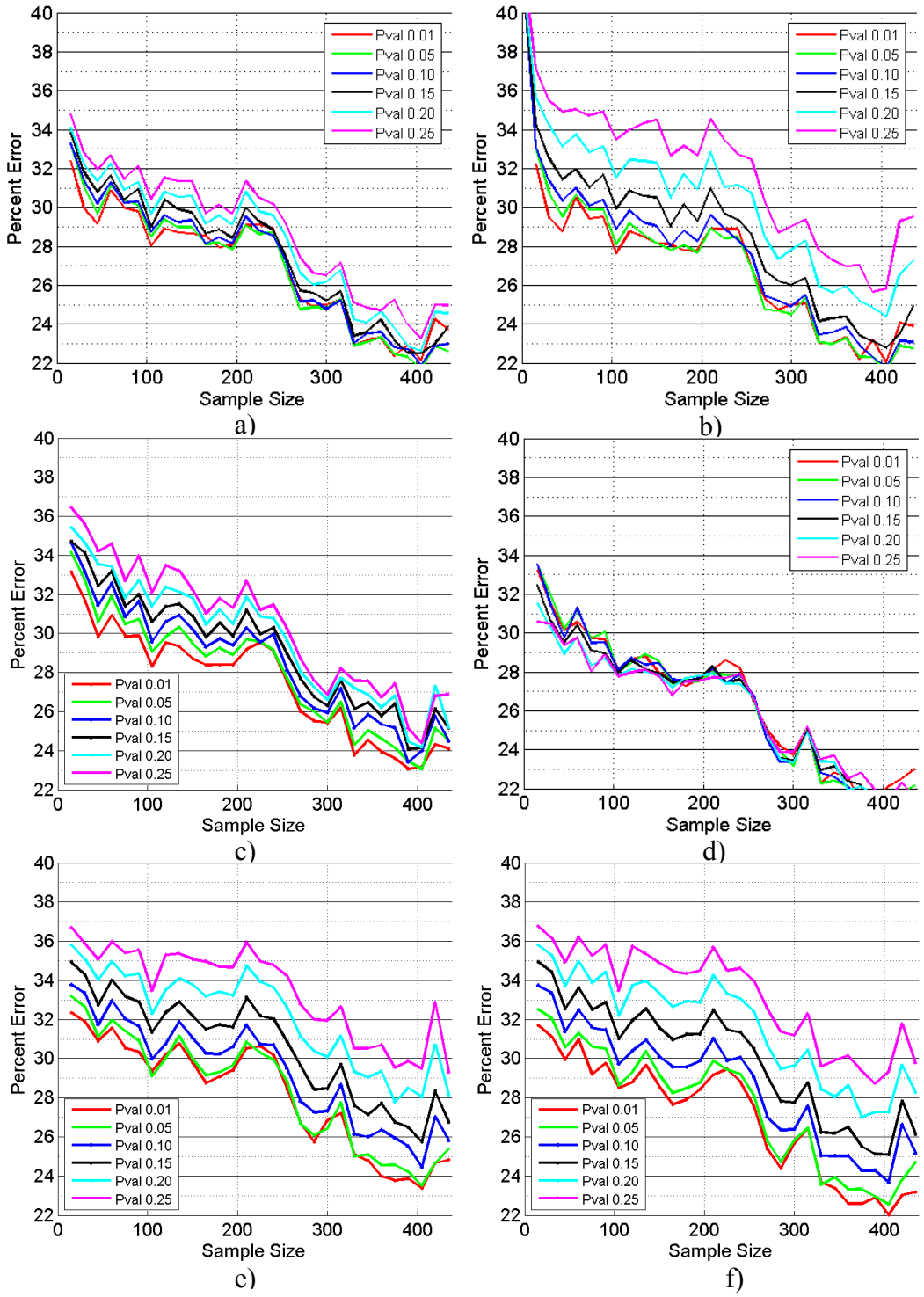



**Figure 140. The impact of varying cumulative variance and penter parameters on the classification of DF data; a) STP with SVM linear kernel; b) STP with SVM RBF kernel; c) STP with LDA; d) STP with QDA; e) STP with kNN (k=1); f) STP with kNN (k=3).**

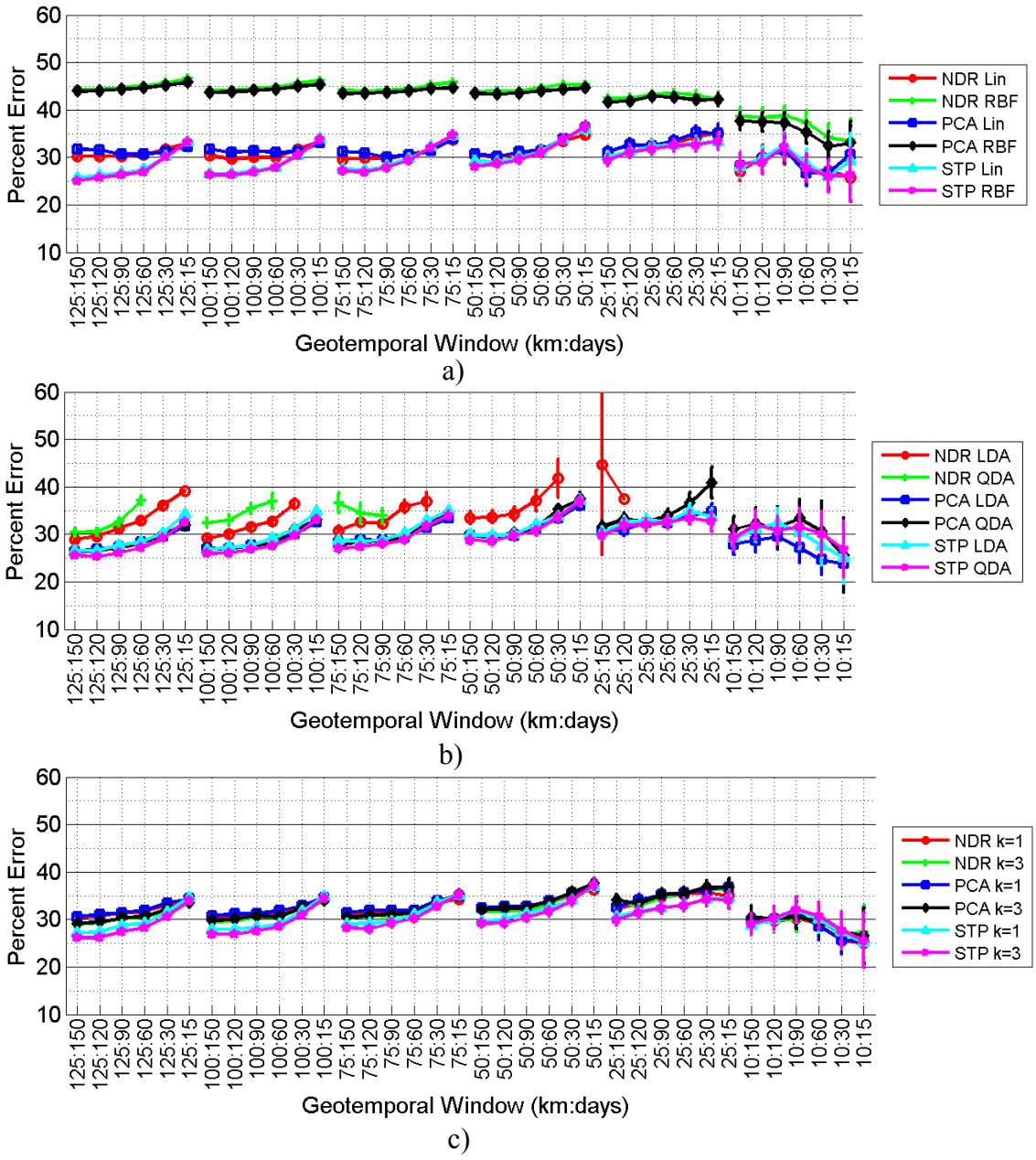
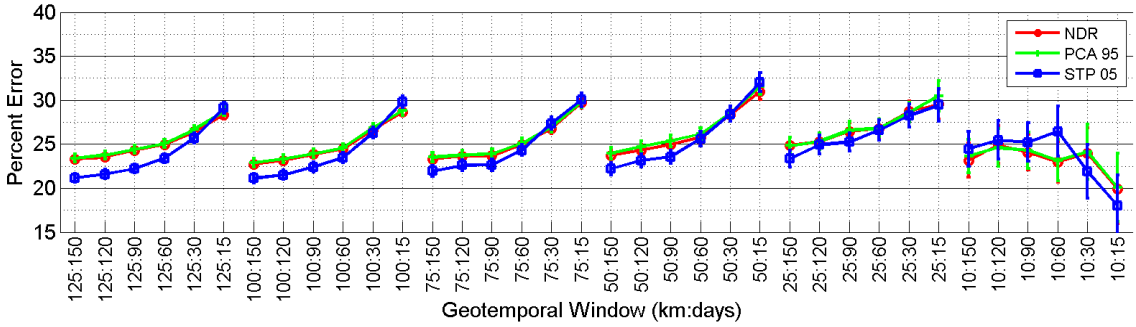
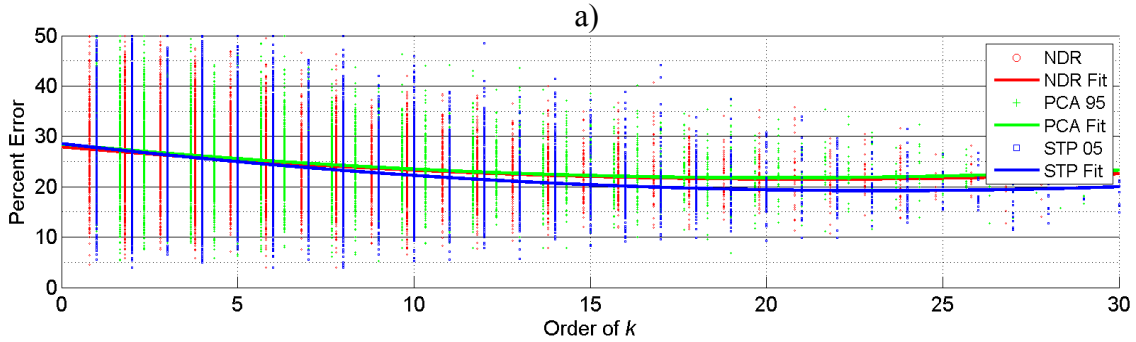


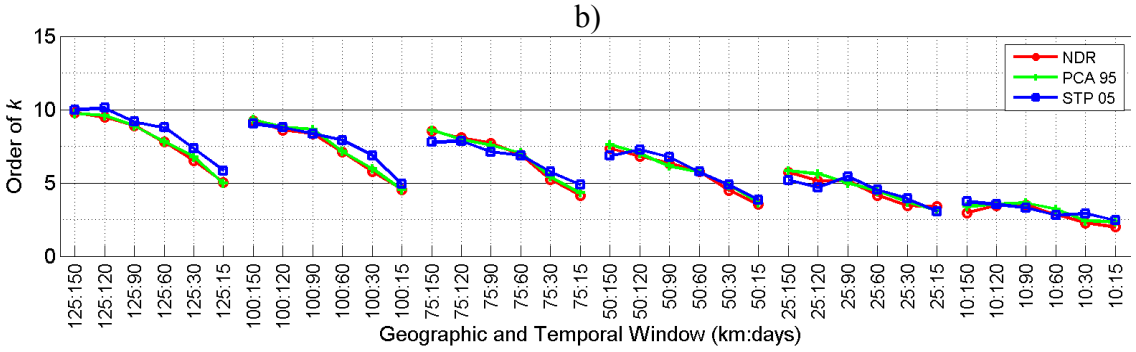

**Figure 141. The impact of dimensionality reduction using PCA (cumulative variance <= 95%) and STP (*penter* <= 0.05) on classification accuracy of DF events when using combined geographic and temporal windows; a) DF SVM; b) DF DA; c) DF kNN.**



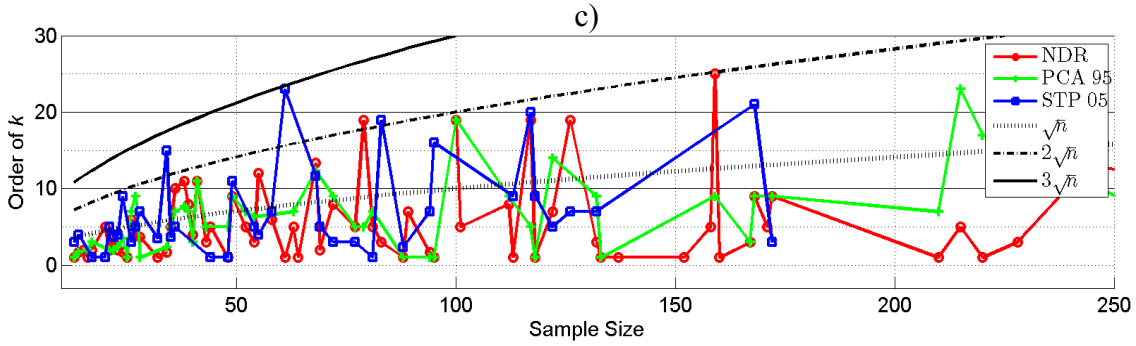

a)




b)




c)




d)

**Figure 142. The impact of dynamically varying the order of *k* on the classification of IED events; a) Classification accuracy under combined geographic and temporal constraints using varying *k*; b) Classification accuracy at various values of *k; c)* Mean order of *k* at various window sizes; d) Mean order of *k* at various sample sizes.**

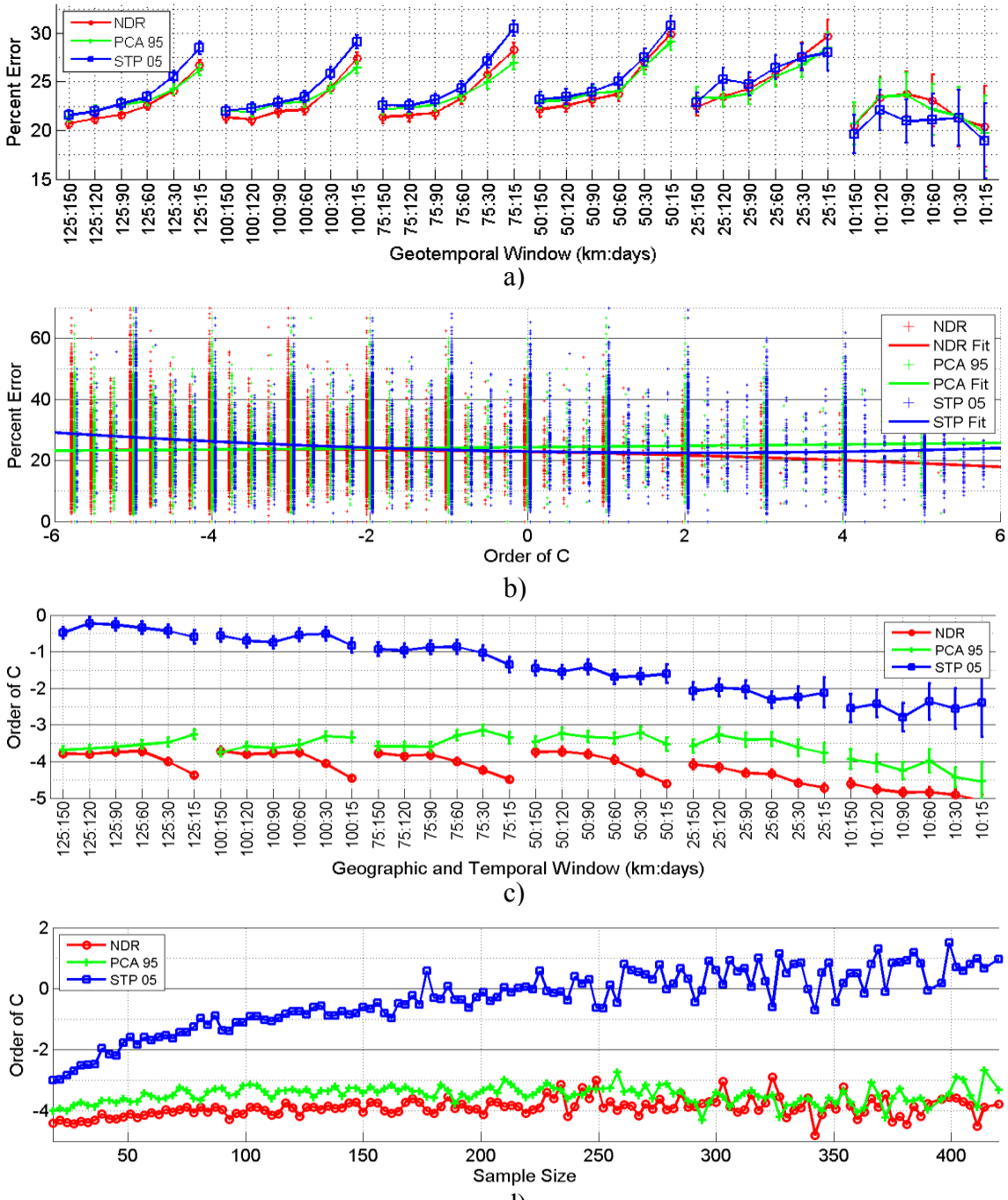



**Figure 143. The impact of dynamically varying the box constraint C on the classification of DF events using SVM with a linear kernel; a) Classification accuracy under combined geographic and temporal constraints using varying *C*; b) Classification accuracy at various values of *C*; c) Mean order of *C* at various window sizes; d) Mean order of *C* at various sample sizes.**

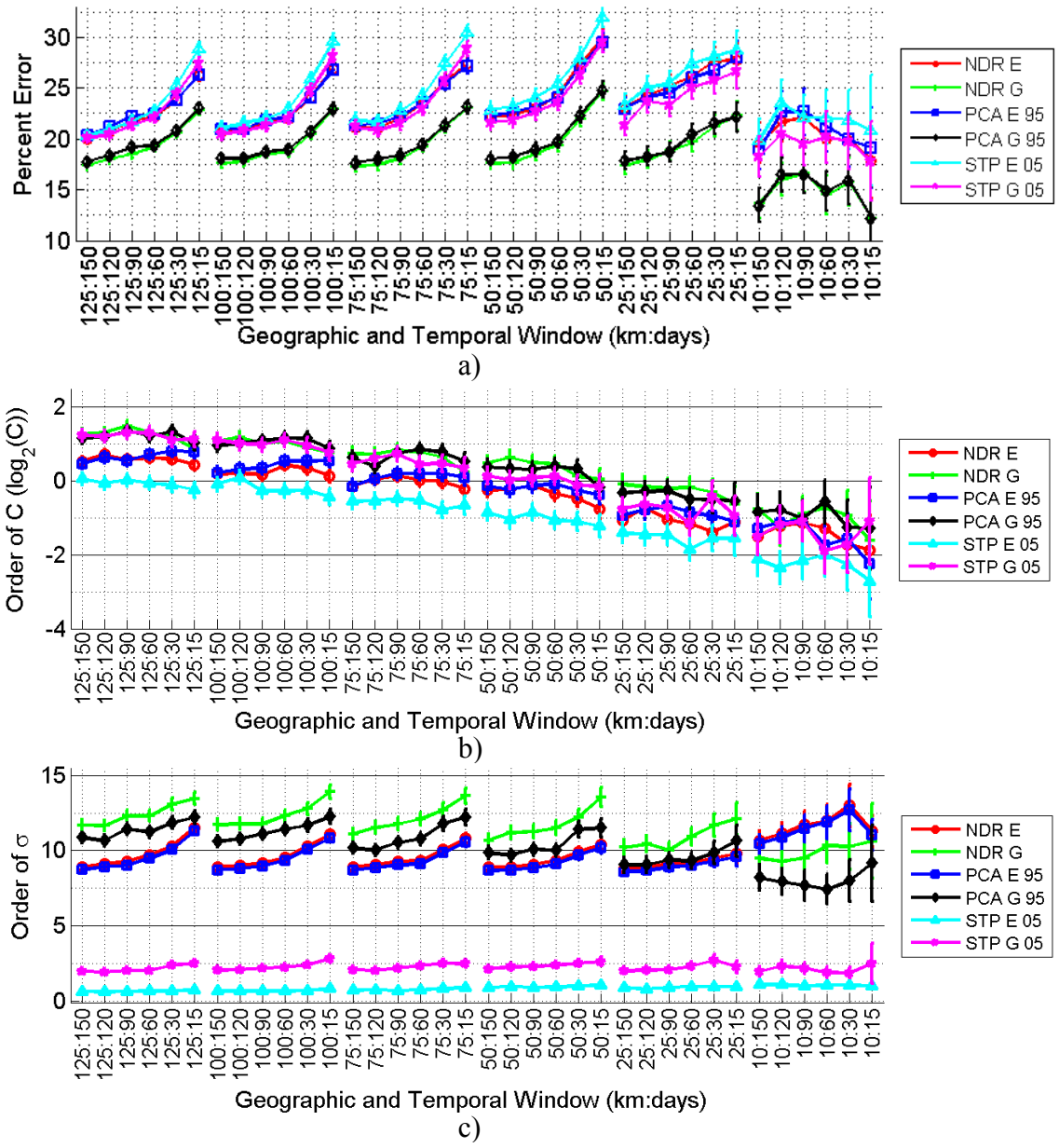



**Figure 144. The impact of dynamically varying the box constraint *C* and the RBF shape parameter *σ* on the classification of DF events using SVM with an RBF kernel; a) Classification accuracy under combined geographic and temporal constraints using varying *C* and *σ*; b) Mean order of *C* at various window sizes; c) Mean order of *σ* at various window sizes.**

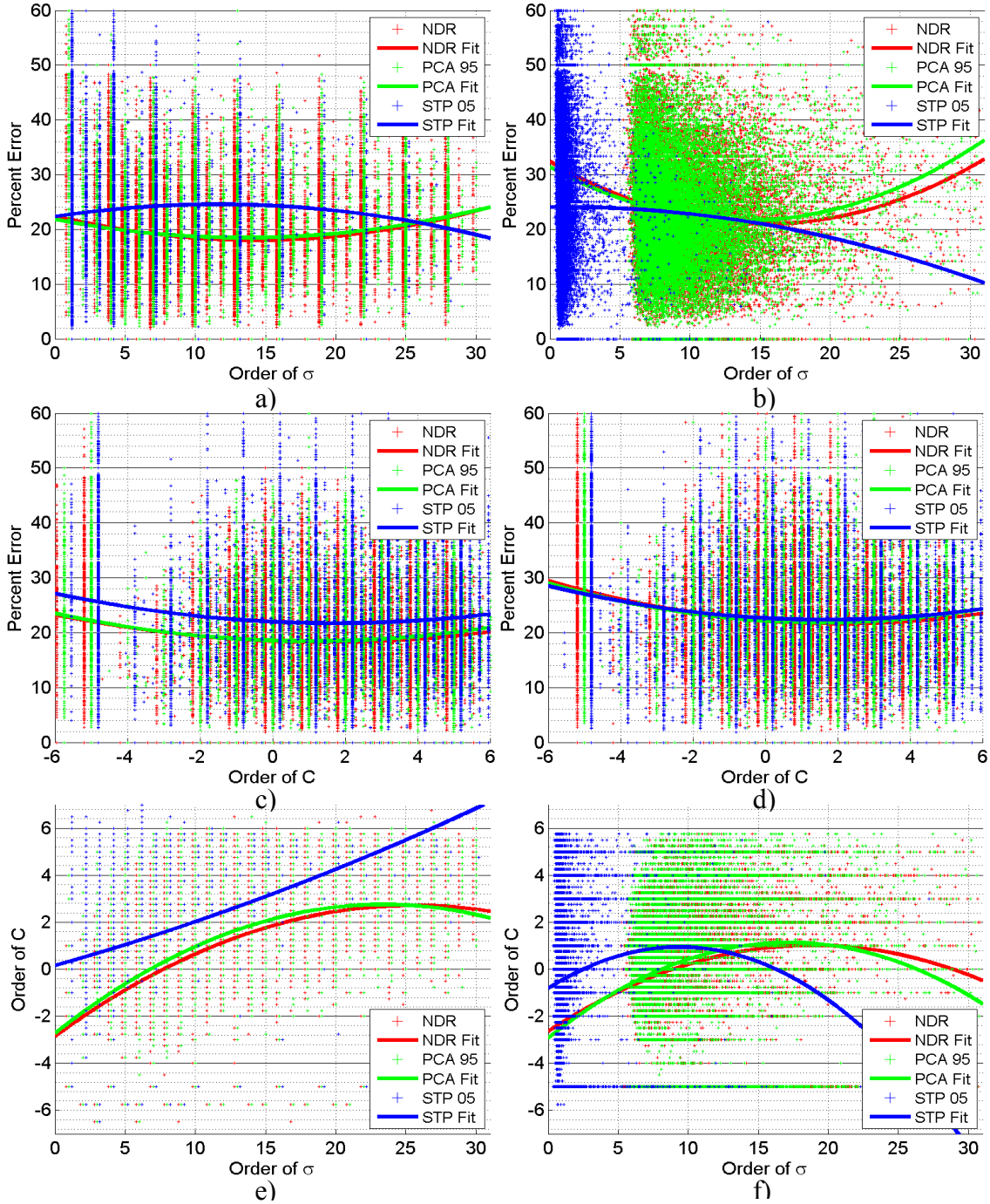



**Figure 145. For DF events, a) Impact of the order of σ found using grid search on classification error; b) Impact of the order of σ found using estimation on classification error; c) Impact of the order of C found using grid search on classification error; d) Impact of the order of C found using estimation on classification error; e) Relationship between C and σ found using grid search; f) Relationship between C and σ found using estimation.**

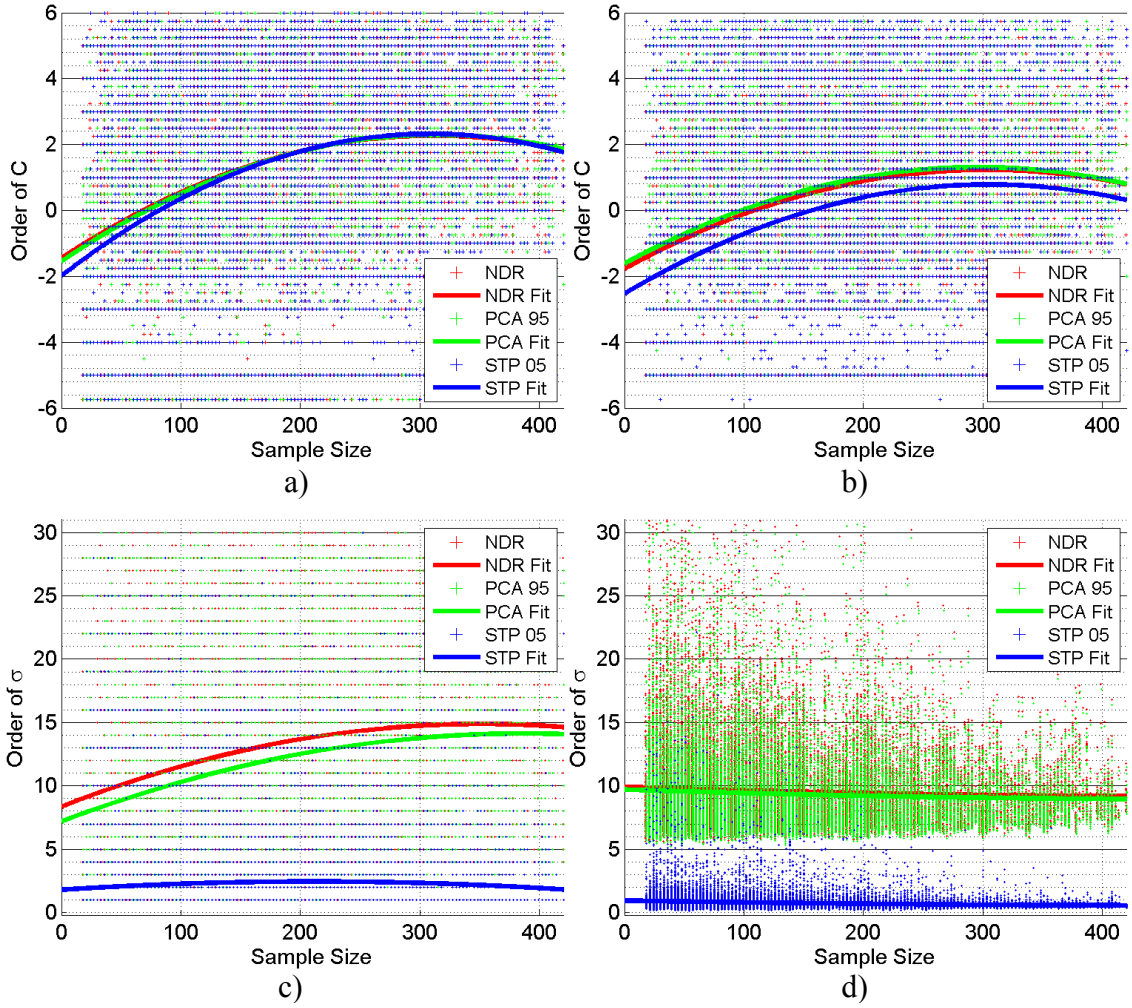



**Figure 146. For DF events, a) Impact of sample size on the order of *C* found using grid search; b) Impact of sample size on the order of *C* found using estimation; c) Impact of sample size on the order of *σ* found using grid search; d) Impact of sample size on the order of *σ* found using estimation.**

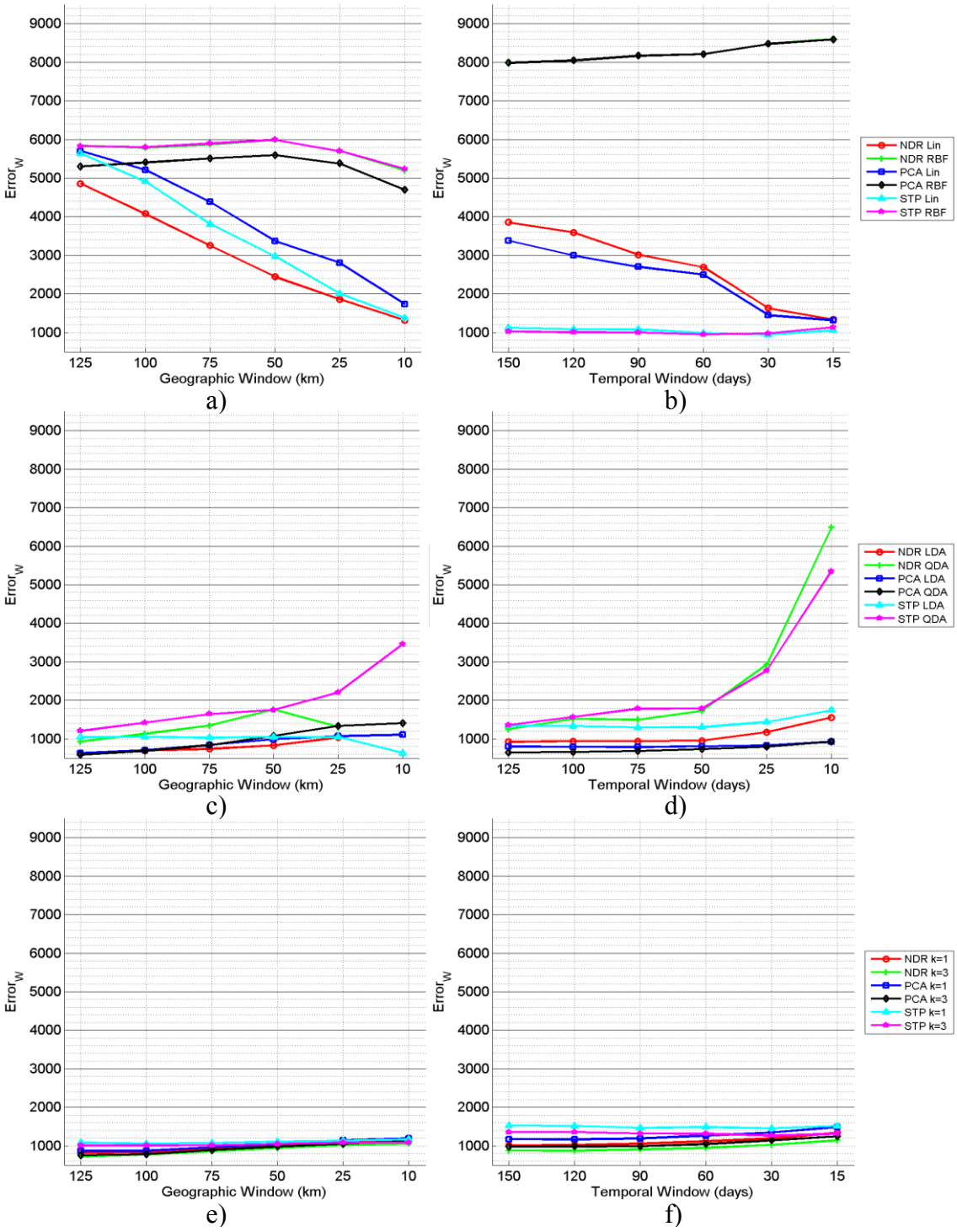



**Figure 147. Weighted cost from learning machines using default parameters on DF data; a) SVM with geographic constraints; b) SVM with temporal constraints; c) DA with geographic constraints; d) DA with temporal constraints; e) kNN with geographic constraints; f) kNN with temporal constraints.**

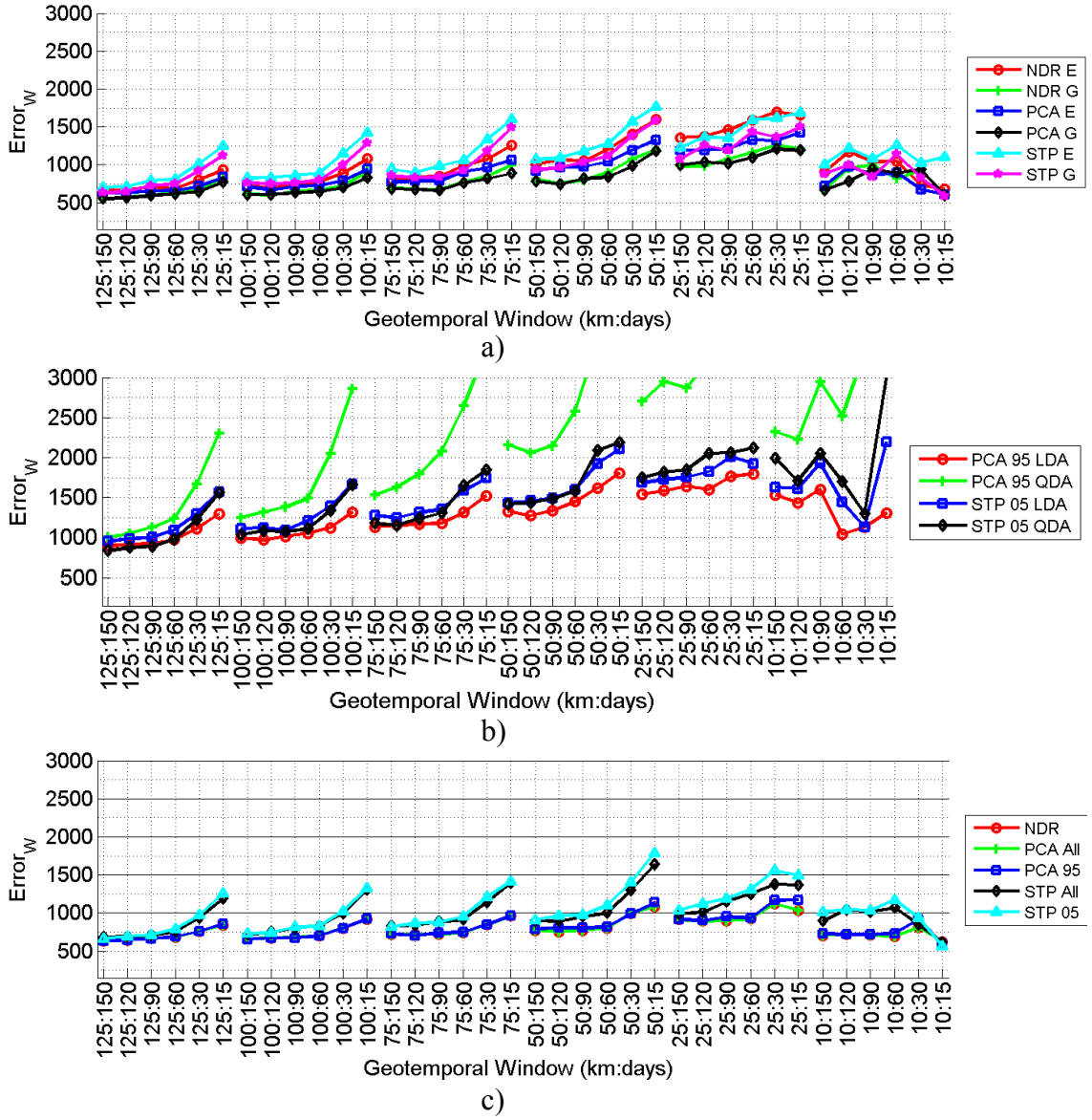



**Figure 148. Weighted cost from learning machines using optimal parameters on DF event data; a) SVM with geotemporal constraints; b) DA with geotemporal constraints; c) kNN with geotemporal constraints.**

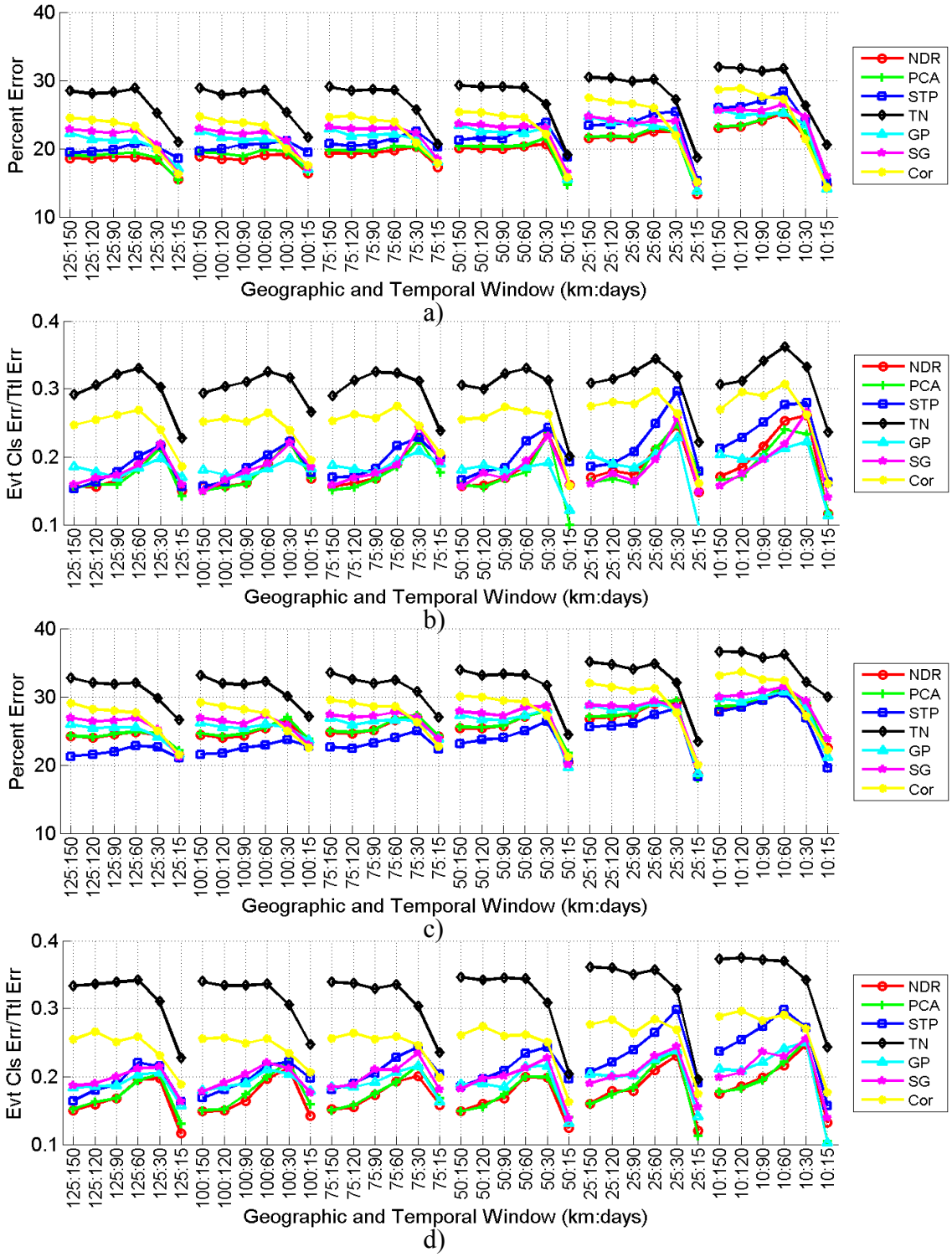



**Figure 149. Subset selection from DF data using experts; a) IED SVM; b) IED SVM classification error analysis; c) IED kNN; d) IED kNN classification error analysis.**

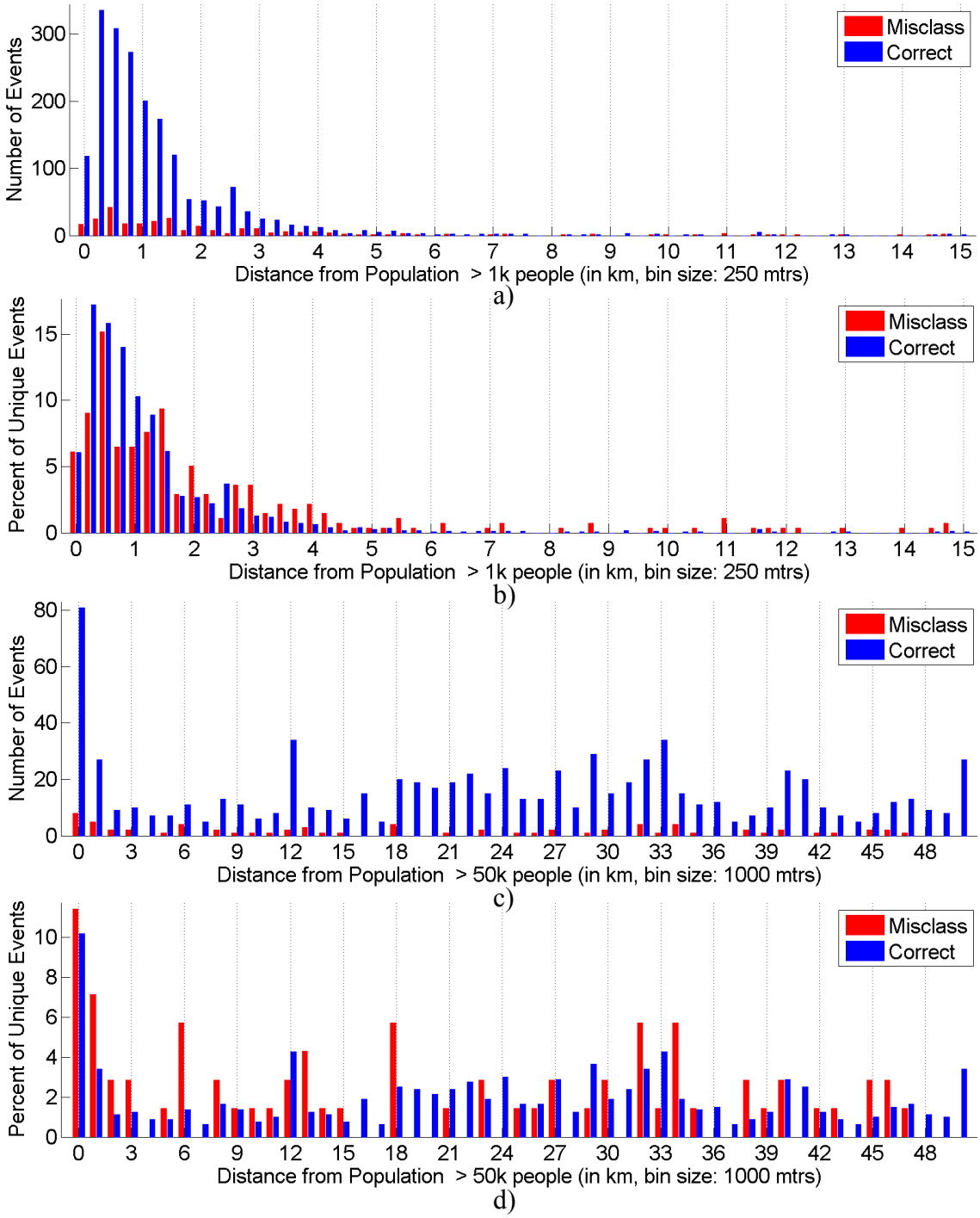



**Figure 150. IED event classification errors by distance from populated areas; a) error count by distance from populated areas with >1000 people; b) DF event classification error ratio over time of day using a sliding window; c) IED event classification errors by time of day and classifier type; d) DF event classification errors by time of day and classifier type.**